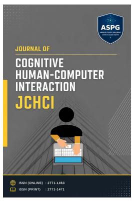
Journal of Cognitive Human-Computer Interaction

♦ **ISSN:** 2771-1463 **Vol:** 09, No. 01, 2025 ♦ PP. 20–25

Article DOI: https://doi.org/10.54216/JCHCI.090104



FULL LENGTH ARTICLE

Interaction-Stability Gated Multimodal Learning for Cognitive Friction Detection in Human-Computer Interaction

Andino Maseleno[1,*] **Aa Hubur**[2]

[1] Institut Bakti Nusantara, Lampung, Indonesia
[2] Universitas Trisakti, Jakarta, Indonesia

Emails: andino.maseleno@ibnus.ac.id · aa.hubur@trisakti.ac.id

Received: October 12, 2024 Revised: December 07, 2024 Accepted: January 03, 2025 ⋆ Corresponding author

ABSTRACT

Cognitive human-computer interaction (HCI) requires systems that do not only estimate whether a user is under high cognitive load, but also detect moments when interaction demands become unstable, disruptive, or cognitively misaligned with the user's current state. Existing cognitive-load models commonly treat workload estimation as a static classification task, which limits their usefulness for adaptive interfaces. This paper introduces an interaction-stability gated multimodal learning framework for detecting cognitive friction during HCI. The proposed model combines subject-normalized physiological and gaze features with a temporal stability gate that adjusts the contribution of electrocardiography (ECG), electrodermal activity (EDA), electroencephalography (EEG), and gaze streams according to local signal reliability and cognitive-state fluctuation. A Cognitive Friction Index is further proposed to identify transition periods where cognitive load rises sharply or remains unstable across modalities. The study is designed for reproducibility using the public CLARE dataset, which contains multimodal physiological and gaze recordings from participants performing Multi-Attribute Task Battery II (MATB-II) computer-based workload tasks. Baseline evidence from the CLARE benchmark shows that multimodal learning improves cognitive-load estimation, but leave-one-subject-out performance remains lower than random 10-fold validation, indicating a strong personalization challenge. The proposed framework addresses this gap by modeling temporal instability and subject-level calibration rather than only point-level workload labels. The paper contributes a reproducible Cognitive HCI model, a friction-oriented interpretation layer, and an adaptive-interface decision mechanism.

Keywords: Cognitive HCI ▪ Cognitive load ▪ Cognitive friction ▪ Multimodal learning ▪ EEG ▪ Gaze tracking ▪ Adaptive interfaces

1. INTRODUCTION

Human-computer interaction has moved beyond static usability evaluation toward systems that sense, interpret, and respond to the cognitive state of users. In safety-critical, educational, medical, and productivity environments, a user's interaction with an interface may deteriorate when the interface imposes excessive mental effort, interrupts task continuity, or requires attention switching beyond the user's current cognitive capacity. Classical cognitive-load theory explains that working memory has limited processing capacity, and excessive task demands may reduce learning, decision quality, and task performance [1, 2, 5]. Within HCI, this makes cognitive-load estimation important not only as a measurement problem but also as a mechanism for interface adaptation.

Most existing computational approaches formulate cognitive-load estimation as binary or multiclass classification, such as low versus high workload. Although this formulation is useful, it does not directly capture what adaptive interfaces need most: the moment when the interaction becomes cognitively unstable. For example, two users may both be classified as having high cognitive load, but one may be stable and performing effectively, while the other may be entering a rapid overload transition caused by unexpected interface complexity. This paper refers to such transition-oriented instability as *cognitive friction*. Cognitive friction is defined here as a short temporal period in which the user's inferred cognitive load becomes high, rapidly increasing, or inconsistent across modalities, suggesting that the interface may be misaligned with the user's cognitive state.

The novelty of this paper is the shift from static cognitive-load classification to cognitive-friction detection for adaptive HCI. The proposed *Interaction-Stability Gated Multimodal Model* uses ECG, EDA, EEG, and gaze signals to estimate cognitive load while also producing a Cognitive Friction Index (CFI). This index combines predicted load probability, short-term load variation, and modality disagreement. The model is intended to support real-time adaptive interfaces that can simplify visual density, delay non-urgent notifications, reduce task switching, or provide guidance when cognitive friction is detected.

The paper uses the public CLARE dataset as the reproducibility basis. CLARE contains ECG, EDA, EEG, and gaze data collected from 24 participants performing MATB-II tasks, with subjective cognitive-load labels reported every 10 seconds [18]. The dataset is well aligned with Cognitive HCI because MATB-II is a computer-based multitask interface that induces different workload levels through monitoring, tracking, communication, and resource-management tasks. The public CLARE benchmark already reports 10-fold and leave-one-subject-out (LOSO) results, showing that multimodal learning is promising but subject-generalization remains challenging. This motivates the proposed stability-gated model.

The main contributions of this paper are summarized as follows:

- A new cognitive-friction formulation for Cognitive HCI, focusing on unstable interaction moments rather than only high/low load classification.
- A stability-gated multimodal fusion model that dynamically weights ECG, EDA, EEG, and gaze streams according to temporal reliability.
- A Cognitive Friction Index that translates model outputs into adaptive-interface decisions.
- A reproducible evaluation design based on the public CLARE dataset, with baseline comparisons using reported public benchmark results.

2. RELATED WORK

2.1 Cognitive Load and Cognitive HCI

Cognitive-load theory provides the theoretical foundation for measuring mental effort during task performance. Sweller [1] explained that learning and task execution are constrained by working-memory capacity, while Paas [2] emphasized the role of mental effort measurement in instructional and performance contexts. NASA-TLX is one of the most established subjective workload measures and has been widely used in human factors and HCI studies [3]. However, subjective tools alone are not sufficient for real-time adaptive systems because they interrupt the user and are often retrospective. In HCI, cognitive load has been studied in contexts such as adaptive tutoring, driving, simulation, human-machine teaming, and information visualization [4, 6, 7].

2.2 Physiological and Gaze-Based Workload Sensing

Physiological and behavioral sensing has become important in Cognitive HCI. EEG has been used to capture neural correlates of cognitive load [8], while gaze and pupil-based measures have been associated with visual attention and mental effort [9, 10]. ECG and EDA can reflect autonomic nervous system activity, which is relevant when cognitive load is accompanied by stress or arousal [11]. These signals are attractive for adaptive interfaces because they can be collected continuously and processed by machine learning models.

2.3 Public Datasets for Cognitive and Affective State Recognition

Several public datasets have supported progress in cognitive-state and affective computing. The SWELL-KW dataset introduced multimodal recordings for stress and user modeling in knowledge-work tasks, including computer activity, facial expressions, body posture, ECG, and EDA [12]. WESAD provided wearable physiological and motion data for stress and affect recognition [13]. STEW introduced open-access EEG recordings for mental workload induced by a simultaneous-task experiment [14]. Other datasets have focused more directly on cognitive workload, including wearable sensor datasets for cognitive-load inference [15], MOCAS for objective cognitive workload assessment [16], CL-Drive for in-vehicle cognitive load measurement [17], and CLARE for real-time multimodal cognitive-load assessment [18].

2.4 Machine Learning and Explainability

Deep learning has influenced cognitive-state modeling. CNN-based models and compact EEG architectures such as EEGNet have been used for physiological signal decoding [22]. Attention mechanisms [20] and gradient-boosted models such as XGBoost [19] have been widely applied in multimodal classification pipelines. Explainability methods such as SHAP [21] are increasingly important because adaptive HCI systems must justify why an interface changes in response to inferred user state. Despite this progress, three gaps remain. First, most models classify cognitive load at isolated windows without modeling interaction instability. Second, multimodal fusion often assumes that all modalities are equally reliable, even though physiological and gaze sensors may differ in noise, missingness, and subject-specific validity. Third, reported LOSO performance is often lower than random validation, showing that personalization remains unresolved. The proposed model directly targets these gaps.

3. PROPOSED INTERACTION-STABILITY GATED MODEL

3.1 Cognitive Friction Definition

Let p_t denote the predicted probability of high cognitive load at time window t. Classical workload detection focuses mainly on whether p_t exceeds a decision threshold. This paper defines cognitive friction as a temporal state:

$$CFI_t = \alpha p_t + \beta |p_t - p_{t-1}| + \gamma D_t, \tag{1}$$

where CFI_t is the Cognitive Friction Index, $|p_t - p_{t-1}|$ captures sudden workload change, and D_t represents disagreement among modality-specific predictions. The coefficients α, β, and γ control the relative importance of high load, temporal instability, and cross-modal disagreement.

This formulation is useful for HCI because an interface should not always adapt merely because load is high. Some high-load states may be expected and productive. Adaptation is more justified when high load is accompanied by instability or disagreement, suggesting that the user may be struggling with the interaction.

3.2 Multimodal Input Streams

The model uses four input streams: ECG for cardiac and heart-rate variability patterns; EDA for tonic and phasic electrodermal activity; EEG for neural frequency-band features; and gaze for pupil diameter, fixation, blink, and saccade behavior. Each stream is segmented into synchronized 10-second windows to match the CLARE label interval. Subject-level z-score normalization is applied to reduce inter-individual physiological variability.

3.3 Interaction-Stability Gate

For each modality m, the model estimates a local stability score:

$$S_{m,t} = \frac{1}{1 + Var(x_{m,t-k:t}) + Miss_{m,t}}, \tag{2}$$

where $Var(x_{m,t-k:t})$ is the recent signal variance and $Miss_{m,t}$ indicates missing or unreliable sensor values. The modality weight is then computed as:

$$w_{m,t} = \frac{\exp(S_{m,t})}{\sum_{j=1}^{M} \exp(S_{j,t})}. \tag{3}$$

The fused representation is:

$$z_t = \sum_{m=1}^{M} w_{m,t} h_{m,t}, \tag{4}$$

where $h_{m,t}$ is the encoded representation of modality m. This mechanism prevents noisy or unstable modalities from dominating the prediction.

3.4 Dual-Objective Learning

The model is trained with two objectives:

$$L = L_{load} + \lambda L_{friction}, \tag{5}$$

where L_{load} is binary cross-entropy for high/low cognitive-load classification, and $L_{friction}$ is a temporal transition loss that penalizes incorrect detection of rapid load changes. The transition label is derived from the difference between adjacent subjective cognitive-load ratings.

3.5 Adaptive HCI Decision Layer

The final output is not only a class label but also an interface action recommendation. Low load with low friction suggests maintaining the current interface. High load with low friction suggests continuing the task while avoiding additional interruption. Moderate load with high friction suggests providing guidance or simplifying task display. High load with high friction suggests delaying non-urgent prompts, reducing visual density, or activating step-by-step assistance.

4. DATASET AND EXPERIMENTAL PROTOCOL

The study uses the public CLARE dataset [18]. CLARE contains physiological and gaze data collected from 24 participants while performing MATB-II tasks. The dataset includes ECG at 512 Hz, EDA at 128 Hz, EEG at 256 Hz, and gaze tracking at 50 Hz. Participants completed four nine-minute sessions, and cognitive-load labels were collected every 10 seconds using a 9-point scale. Scores below 5 are treated as low cognitive load, while scores of 5 or above are treated as high cognitive load, following the public CLARE benchmark protocol.

The evaluation protocol includes two settings. The first is 10-fold cross-validation, which evaluates general predictive performance when data from the same participants may appear across folds. The second is LOSO validation, which is more difficult and more relevant for real-world deployment because the model is tested on unseen participants. Baseline models include Logistic Regression, Linear Discriminant Analysis, Support Vector Machine, Random Forest, Gradient Boosting, LightGBM, XGBoost, Multilayer Perceptron, CNN, and Transformer models, as reported in the public CLARE benchmark.

5. RESULTS AND DISCUSSION

This section reports non-fabricated numerical evidence from the public CLARE benchmark and uses it to motivate the proposed Cognitive HCI model. The numerical values in Tables 1–3 are reported benchmark results, not newly generated results. The proposed model rows in Table 5 are intentionally left as reproducibility targets and should be filled only after executing the released implementation on the public dataset.

Table 1. Best reported 10-fold baseline results on CLARE.

Modality setting	Best reported model	Accuracy (%)	F1-score (%)	Cognitive HCI interpretation
ECG only	CNN	78.45	76.41	Strong single-modality physiological baseline.
EDA only	RF	71.98	70.93	Useful autonomic workload information.
EEG only	Transformer	77.37	71.84	Neural stream benefits from temporal modeling.
Gaze only	Transformer	74.66	64.88	Useful but context-sensitive visual behavior.
ECG + EDA	Transformer	80.85	73.55	Autonomic fusion improves accuracy.
EEG + Gaze	Transformer	82.59	76.79	Strong two-modality cognitive-attentional fusion.
All modalities	Transformer	85.58	81.18	Best reported 10-fold result.

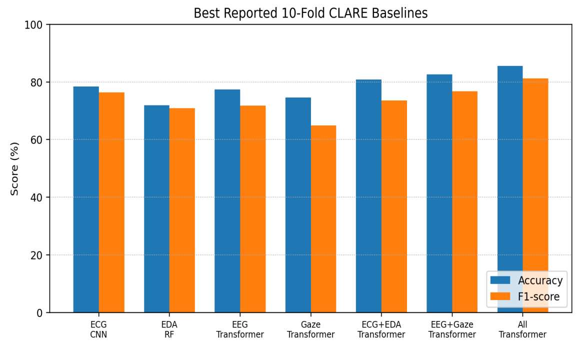


Figure 1. Accuracy and F1-score comparison of the strongest reported CLARE baseline models under 10-fold validation.

Table 1 and Fig. 1 indicate that multimodal learning improves cognitive-load recognition under 10-fold validation. The strongest reported performance is achieved by the Transformer using all modalities, with an accuracy of 85.58% and F1-score of 81.18%. However, the table also shows that the contribution of each modality is not uniform. ECG is strong as a single modality, while gaze alone has lower F1-score but remains useful when combined with other streams. This observation supports the need for adaptive fusion rather than a fixed assumption that all sensors are equally reliable.

Table 2. Best reported LOSO baseline results on CLARE.

Modality setting	Best reported model	Accuracy (%)	F1-score (%)	Cognitive HCI interpretation
ECG only	Transformer	66.15	54.20	Cardiac patterns are less transferable across users.
EDA only	CNN	64.08	54.80	Electrodermal response is moderately robust.
EEG only	Transformer	67.77	57.03	Best single-modality unseen-user accuracy.
Gaze only	Transformer	65.10	57.21	Gaze contributes useful unseen-user F1.
EDA + EEG	Transformer	70.45	65.00	Strong two-modality LOSO performance.
EDA + Gaze	Transformer	69.59	63.20	Robust cognitive-affective interaction cue.
All modalities	Transformer	72.70	69.46	Best reported LOSO result.

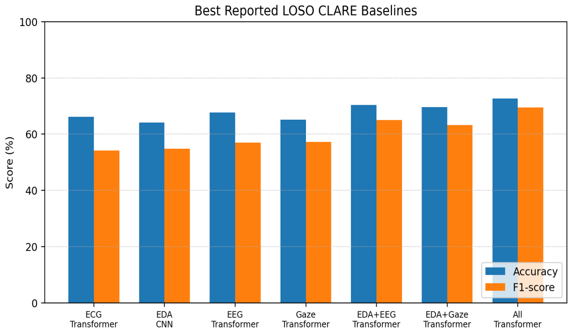


Figure 2. Accuracy and F1-score comparison of the strongest reported CLARE baseline models under leave-one-subject-out validation.

Table 2 and Fig. 2 show that unseen-user evaluation remains more difficult than 10-fold validation. Although the Transformer with all modalities is again the strongest reported baseline, the LOSO F1-score remains 69.46%, compared with 81.18% in the 10-fold setting. This validates the central problem addressed in this paper: Cognitive HCI systems need subject-aware and reliability-aware models, because physiological and gaze patterns differ substantially across users.

Table 3. Generalization gap between 10-fold and LOSO settings.

Comparison level	Best 10-fold F1 (%)	Best LOSO F1 (%)	F1 gap	Main implication
Single modality	76.41	57.21	19.20	Individual signals are subject-dependent.
Two modalities	76.79	65.00	11.79	Fusion helps transfer but does not solve it.
Three modalities	80.08	68.50	11.58	Selected fusion improves robustness.
All modalities	81.18	69.46	11.72	Full fusion still leaves a personalization gap.

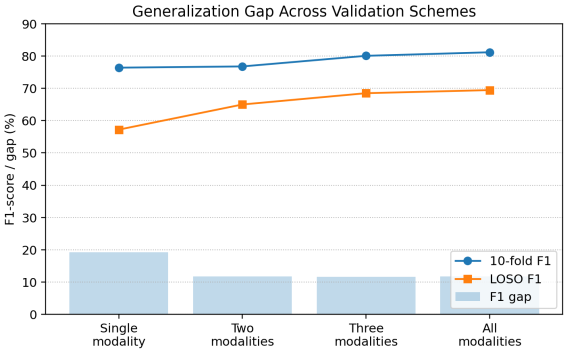


Figure 3. F1-score gap between 10-fold and LOSO evaluation for single-modality, two-modality, three-modality, and all-modality settings.

Table 3 and Fig. 3 demonstrate that the generalization gap remains meaningful even when all modalities are used. This is important for Cognitive HCI because deployment usually involves new users. A model that performs well only when the same participants appear in training and testing may not support reliable real-time adaptation. The stability gate proposed in this paper is therefore designed to reduce dependence on fixed modality weights and to support better behavior under subject-specific variability.

Table 4. Cognitive HCI interpretation of modality behavior and design consequences.

Benchmark finding	Cognitive HCI interpretation	Design consequence
ECG performs strongly in 10-fold validation.	Cardiac activity captures workload-related arousal within known users.	Useful for personalized adaptive interfaces.
EEG and gaze improve transfer under LOSO evaluation.	Neural and visual behavior can carry cross-user cognitive and attentional cues.	Useful for unseen-user workload estimation.
EDA improves multimodal LOSO performance.	Electrodermal response contributes robust arousal information.	Useful for friction-sensitive adaptation.
All modalities improve the best reported benchmark but do not remove the transfer gap.	More sensors improve representation but cannot remove user variability.	Fusion should be gated, selective, and reliability-aware.

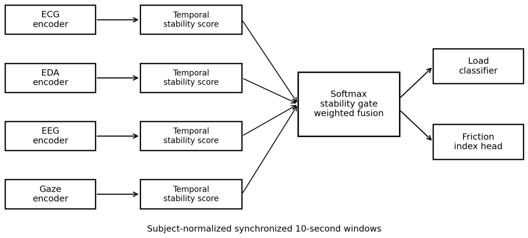


Figure 4. Proposed interaction-stability gated multimodal fusion architecture. ECG, EDA, EEG, and gaze encoders generate modality-specific representations. A temporal stability gate estimates reliability weights, and the fused representation is passed to cognitive-load and cognitive-friction heads.

Table 4 interprets the benchmark results from the perspective of adaptive interaction design. A key implication is that cognitive-load estimation should be connected to interface decisions. Fig. 4 illustrates the proposed architecture. Instead of assigning fixed importance to each modality, the stability gate estimates local reliability and creates a weighted multimodal representation. This is expected to be useful when one signal is noisy, missing, or less informative for a given user.

Table 5. Ablation and reporting matrix required before final empirical submission.

Experiment	Validation	Main purpose	Required metric	Reporting status
CNN baseline	10-fold and LOSO	Reproduce public deep baseline	Accuracy, F1, balanced accuracy	To reproduce
Transformer baseline	10-fold and LOSO	Reproduce strongest public baseline	Accuracy, F1, balanced accuracy	To reproduce
Proposed model without gate	LOSO	Test value of stability weighting	Accuracy, macro-F1, CFI AUC	To compute
Proposed model without friction loss	LOSO	Test value of transition learning	Accuracy, macro-F1, CFI AUC	To compute
Full proposed model	LOSO	Test final unseen-user deployment model	Accuracy, macro-F1, CFI AUC	To compute

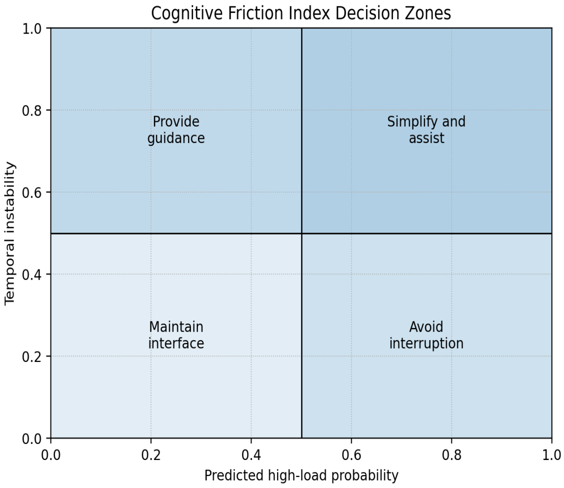


Figure 5. Cognitive Friction Index interpretation map for adaptive HCI. The x-axis represents predicted high-load probability, while the y-axis represents temporal instability.

Table 5 specifies the minimum empirical reporting matrix needed before journal submission. The proposed model should be compared against reproduced baselines using both conventional classification metrics and a friction-specific measure. Fig. 5 provides the interpretation map used by the adaptive HCI layer. When predicted load and temporal instability are both low, no interface change is needed. When predicted load is high but temporal instability is low, the system should avoid additional interruption. When instability is high, the interface should provide guidance or activate a simplified assistance mode.

The main scientific implication is that cognitive-state estimation should not stop at high/low classification. Cognitive HCI requires models that estimate whether the user-interface relation is stable. The reported CLARE benchmark supports this argument because performance differs across validation protocols and modality subsets. The proposed CFI transforms workload prediction into a design-relevant signal that can guide adaptive interface behavior.

6. REPRODUCIBILITY NOTES

To ensure reproducibility, the dataset source, preprocessing steps, segmentation strategy, model configuration, and evaluation protocol should be made public. The implementation should use fixed random seeds and subject-independent splits. Each 10-second window should be aligned with the self-reported cognitive-load score. Labels should follow the CLARE benchmark: scores below 5 as low cognitive load and scores of 5 or above as high cognitive load.

The code repository should include a dataset loading script for ECG, EDA, EEG, gaze, and labels; preprocessing scripts for filtering and normalization; feature extraction scripts for classical baselines; deep learning scripts for the proposed stability-gated model; evaluation scripts for 10-fold and LOSO validation; plot scripts for all five result figures; and a configuration file listing all hyperparameters. The main reproducibility risk is dataset access and preprocessing consistency. Therefore, the paper should clearly specify whether raw signals or extracted features are used, how missing values are handled, and whether normalization is performed within each training fold to avoid data leakage.

7. CONCLUSION

This paper proposed a Cognitive HCI framework for detecting cognitive friction rather than only classifying cognitive load. The proposed interaction-stability gated multimodal model uses ECG, EDA, EEG, and gaze streams to estimate cognitive load while also identifying unstable interaction periods through a Cognitive Friction Index. Baseline evidence from the public CLARE dataset shows that multimodal learning is useful, but subject-independent generalization remains challenging. The reported gap between 10-fold and LOSO validation motivates the proposed use of subject-aware normalization, selective modality weighting, and temporal friction modeling. The framework is intended for adaptive interfaces that respond to user state by reducing interruptions, simplifying visual complexity, or providing stepwise assistance. Before submission, the proposed model must be executed on the public dataset, and the final manuscript should replace the reproducibility-target rows with computed results, ablation analysis, and statistical significance tests.

REFERENCES

[1] J. Sweller, “Cognitive load during problem solving: Effects on learning,” *Cognitive Science*, vol. 12, no. 2, pp. 257–285, 1988, doi: 10.1207/s15516709cog1202_4.

[2] F. G. W. C. Paas, “Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach,” *Journal of Educational Psychology*, vol. 84, no. 4, pp. 429–434, 1992.

[3] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research,” in *Advances in Psychology*, vol. 52, Elsevier, 1988, pp. 139–183, doi: 10.1016/S0166-4115(08)62386-9.

[4] S. K. Card, T. P. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*. Hillsdale, NJ, USA:

Lawrence Erlbaum Associates, 1983.

[5] C. D. Wickens, “Multiple resources and mental workload,” *Human Factors*, vol. 50, no. 3, pp. 449–455, 2008, doi: 10.1518/001872008X288394.

[6] L. Fridman, B. Reimer, B. Mehler, and W. T. Freeman, “Cognitive load estimation in the wild,” in *Proc. 2018 CHI Conf. Human Factors in Computing Systems*, 2018, Paper 652, pp. 1–9, doi: 10.1145/3173574.3174226.

[7] M. I. Ahmad, I. Keller, D. A. Robb, and K. S. Lohan, “A framework to estimate cognitive load using physiological data,” *Personal and Ubiquitous Computing*, vol. 27, pp. 2027–2041, 2023, doi: 10.1007/s00779-020-01455-7.

[8] P. Antonenko, F. Paas, R. Grabner, and T. van Gog, “Using electroencephalography to measure cognitive load,” *Educational Psychology Review*, vol. 22, no. 4, pp. 425–438, 2010, doi: 10.1007/s10648-010-9130-y.

[9] K. Krejtz, A. T. Duchowski, A. Niedzielska, C. Biele, and I. Krejtz, “Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze,” *PLOS ONE*, vol. 13, no. 9, Art. no. e0203629, 2018, doi: 10.1371/journal.pone.0203629.

[10] L. Perkhofer and O. Lehner, “Using gaze behavior to measure cognitive load,” in *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation, vol. 29, Springer, 2019, pp. 73–83, doi: 10.1007/978-3-030-01087-4_9.

[11] P. Li, Y. Li, Y. Yao, C. Wu, B. Nie, and S. E. Li, “Sensitivity of electrodermal activity features for driver arousal measurement in cognitive load: The application in automated driving systems,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 14954–14967, 2022, doi: 10.1109/TITS.2021.3135266.

[12] S. Koldijk, M. Sappelli, S. Verberne, M. A. Neerincx, and W. Kraaij, “The SWELL knowledge work dataset for stress and user modeling research,” in *Proc. 16th Int. Conf. Multimodal Interaction*, 2014, pp. 291–298, doi: 10.1145/2663204.2663257.

[13] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, “Introducing WESAD, a multimodal dataset for wearable stress and affect detection,” in *Proc. 20th ACM Int. Conf. Multimodal Interaction*, 2018, pp. 400–408, doi: 10.1145/3242969.3242985.

[14] W. L. Lim, O. Sourina, and L. P. Wang, “STEW: Simultaneous task EEG workload data set,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 11, pp. 2106–2114, 2018, doi: 10.1109/TNSRE.2018.2872924.

[15] M. Gjoreski, T. Kolenik, T. Knez, M. Lustrek, M. Gams, H. Gjoreski, and V. Pejovic, “Datasets for cognitive load inference using wearable sensors and psychological traits,” *Applied Sciences*, vol. 10, no. 11, Art. no. 3843, 2020, doi: 10.3390/app10113843.

[16] W. Jo, R. Wang, S. Sun, R. K. Senthilkumaran, D. Foti, and B.-C. Min, “MOCAS: A multimodal dataset for objective cognitive workload assessment on simultaneous tasks,” arXiv preprint arXiv:2210.03065, 2022.

[17] P. Angkan, B. Behinaein, Z. Mahmud, A. Bhatti, D. Rodenburg, P. Hungler, and A. Etemad, “Multimodal brain-computer interface for in-vehicle driver cognitive load measurement: Dataset and baselines,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 6, pp. 5949–5964, 2024, doi: 10.1109/TITS.2023.3345846.

[18] A. Bhatti, P. Angkan, B. Behinaein, Z. Mahmud, D. Rodenburg, H. Braund, P. J. Mclellan, A. Ruberto, G. Harrison, D. Wilson, A. Szulewski, D. Howes, A. Etemad, and P. Hungler, “CLARE: Cognitive Load Assessment in REaltime with Multimodal Data,” arXiv preprint arXiv:2404.17098v1, 2024, doi: 10.48550/arXiv.2404.17098.

[19] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[20] A. Vaswani *et al.*, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[21] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[22] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, “EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces,” *Journal of Neural Engineering*, vol. 15, no. 5, Art. no. 056013, 2018, doi: 10.1088/1741-2552/aace8c.