CogFormer: Aligned-attention Transformer-based Multi-physiological signals Fusion for Driver Cognitive Load Estimation in Conditional Automated Driving

Ange Wang
Intelligent Transportation Thrust, Systems Hub
The Hong Kong University of Science and Technology (Guangzhou), China
Email: awang324@connect.hkust-gz.edu.cn

Haohan Yang
School of Mechanical and Aerospace Engineering
Nanyang Technological University, Singapore
Email: haohan.yang@ntu.edu.sg

Jiyao Wang
Robotics and Autonomous Systems Thrust, Systems Hub
The Hong Kong University of Science and Technology (Guangzhou), China
Email: jwang297@connect.hkust-gz.edu.cn

Hai Yang
Chair Professor
Department of Civil and Environmental Engineering
The Hong Kong University of Science and Technology, Hong Kong SAR, China
Email: cehyang@ust.hk

Dengbo He
Assistant Professor, Corresponding author
Intelligent Transportation Thrust, Systems Hub
The Hong Kong University of Science and Technology (Guangzhou), China
HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen
Email: dengbohe@hkust-gz.edu.cn

Word Count: 1,693 words + 4 tables/figures

Submitted [November, 19, 2024]

Statement of Significance (Relevance of Research)

This research introduces CogFormer, a decision-level fusion model to accurately estimate driver cognitive load in conditional automated vehicles. By integrating physiological signals with an attention mechanism, CogFormer provides robust and real-time estimation of driver state, surpassing existing models in accuracy. The findings have implications for enhancing safety in SAE Level 3 and above by enabling cognitive load estimation for adaptive support systems, including but not limited to takeover assistance and fatigue monitoring. This research will be relevant to TRB attendees who are interested in driver monitoring, intelligent transportation systems, and smart cabins.

Acknowledgments

This work was supported by the Natural Science Foundation of Guangdong Province of China (2024A1515010392) and partially by the National Natural Science Foundation of China (52202425).

Author Contribution

The authors confirm their contribution to the paper as follows: **Ange Wang**: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft. **Haohan Yang**: Validation, Formal analysis, coding assistance. **Jiyao Wang**: Validation, Formal analysis. **Hai Yang**: Validation, Formal analysis. **Dengbo He**: Formal analysis, Funding acquisition, Methodology, Supervision, Validation, Writing – review & editing.

INTRODUCTION

Human error is recognized as one of the dominating factors in road accidents (*1*). Compared to driving tasks that are visually and manually demanding, the cognitive demanding tasks can be more safety-critical, and thus drivers' performance in these tasks has been widely adopted as key metrics differentiating novice and experienced drivers (*2*). The introduction of infotainment functions in the smart cabin and the prevalence of bring-in smart devices may also increase the task load of drivers.

The high cognitive load in driving has been found to be closely related to driving safety. For example, a high cognitive load may lead to delayed responses to emergency events (*3*), visual tunnel effect (*4*), decreased ability to anticipate hazards (*5*). The introduction of driving automation may be a solution, as a lower overall task load has been observed in vehicles equipped with advanced driving automation systems (ADASs) (*6*). However, some studies found that ADAS can increase drivers' cognitive load due to the additional responsibility to monitor automation (*7*).

Most of the previous cognitive load estimation algorithms were developed for non-automated vehicles, which may not apply to vehicles with ADAS. For example, the study by Wiediartini et al. (*8*) revealed significant differences in fixation duration and pupil diameter across the three task difficulty levels of memory and arithmetic tasks. This discrepancy in physiological measures has also been observed between manual and automated driving.

To the best of our knowledge, we can only identify three studies that focused on high cognitive load estimation in vehicles with driving automation (*9–11*), and most of the existing driver cognitive estimation approaches (including the ones for non-automated vehicles and vehicles with driving automation) have two major limitations:

- Most previous models relied on manual feature extraction of physiological features (*3*, *11*, *12*), for example, heart rate (HR) extracted from electrocardiogram (ECG) (*13*), and respiratory rate extracted from respiratory signals (RESP) (*14*). While satisfactory accuracy has been achieved in previous research, the extraction of these handcrafted low-level features is computational costing and may lead to loss of information in the raw signals.
- Most driver cognitive load estimation studies used classical machine learning models that ignored the temporal-spatial dependency of physiological signals. Only a few studies used Recurrent Neural Networks (RNNs) (*12*) and Long Short-Term Memory (LSTM) networks (*15*) that could consider temporal information for driver cognitive load estimation. However, RNN or LSTM may still neglect the long-distance temporal dependencies present in time series and previous studies did not fuse multiple physiological features, ignoring the spatial dependency among signals.

To address the aforementioned challenges, in this study, we proposed an Attention-Aligned Transformer algorithm for Cognitive (CogFormer) load estimation in vehicles with driving automation by integrating ECG, electrodermal activity (EDA), and RESP signals.

METHODOLOGY

The aligned attention mechanism is the most important component of CogFormer which integrates multiple physiological signals, i.e., ECG, RESP, and EDA signals, by utilizing self-attention and aligned attention. Self-attention captures dependencies within each signal type, while aligned attention aligns and integrates information across different signals. This approach allows the model to weigh and combine the most relevant features from each signal, creating a cohesive and comprehensive representation. The ability of the mechanism to fuse diverse data sources can allow the model to understand and predict physiological patterns by capturing complex interdependencies. Cross-modal attention and self-attention mechanisms are expressed by equations (1) and (2), respectively.

$$f_{ii} = softmax\left(\frac{Q_i {K_i}^T}{\sqrt{d_k}}\right) V_i, \tag{1}$$

$$f_{ij} = softmax\left(\frac{Q_i {K_j}^T}{\sqrt{d_k}}\right) V_j, i, j \in s \; and \; i \neq j. \tag{2}$$

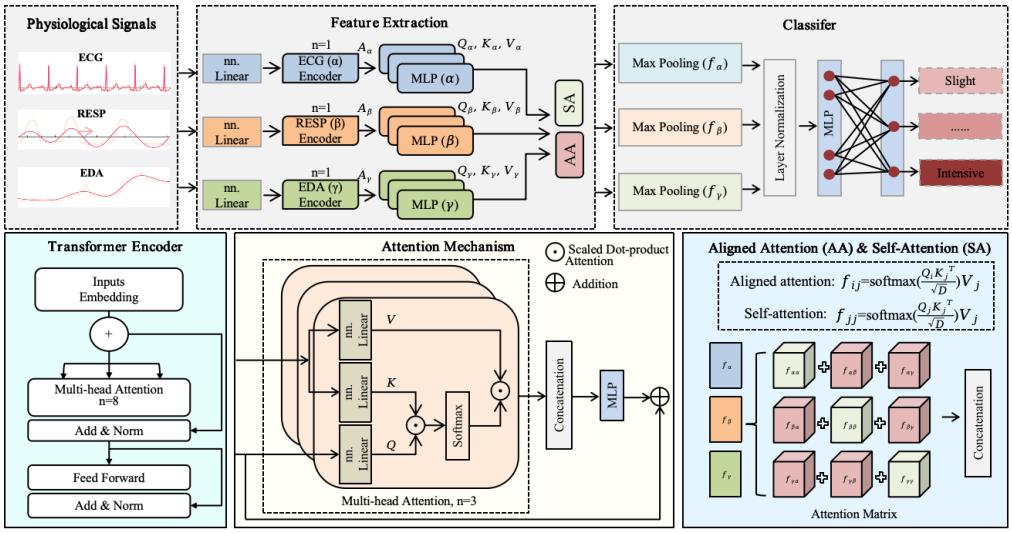

FIGURE 1 Overview of the proposed CogFormer for driver cognitive load detection.

EXPERIMENT AND DATASET

We tested our model on the mathematical and autonomous driving tasks dataset (MADT-D) published by the University of Applied Sciences and Arts of Western Switzerland (*16*). The MADT-D consists of driving data from 90 participants (with two participants deemed invalid). In the experiment, half of the participants were instructed to perform a cognitive task known as the oral digit span counting task, requiring them to verbally count backward from 3,645 in decrements of 2 (labeled as high cognitive load), while the rest half conducted a driving task only (labeled as low cognitive load). To ensure that both load levels appear in the leave-one-out test, we combined one participant performing the non-driving-related tasks (NDRTs) with one participant not performing the NDRT to form a new subject. Consequently, there are a total of 44 new participants. Similarly, in the experiment, the physiological data, i.e., EDA, RESP, and ECG, were collected using sensors by BioPac at a frequency of 1,000Hz in the dataset. The differences in subjective ratings of cognitive load between tasks were validated in Meteier et al. (*9*). We extracted data spanning 10 minutes from the cognitive load phase in the MADT-D dataset.

Apart from our own dataset, to better validate our proposed model, we also constructed a dataset, named CAM-CLD, which focuses on drivers' cognitive states in simulated SAE Level-3 vehicles. The details of the experiment for CAM-CLD are provided below. This study was approved by the Hong Kong University of Science and Technology (HREP-2023-0199).

Participants

In total, 42 drivers (25 males, 17 females, aged 23-53) were recruited, ensuring a balanced age distribution to improve generalizability. All had at least one year of licensed driving experience and were compensated at a rate of 70 RMB per hour.

Cognitive Load Tasks

Three kinds of NDRTs used in CAM-CLD, i.e., the n-back task in which participants needed to recall stimuli presented n positions earlier (0-, 1-, 2-back) (*17*); math calculation task, in

which the participants needed to verbally count backward from 3000 by a step of 3 or 5 (*9*); and spatial memory, in which participants needed to recall the final direction after listening to an audio description of a route (*18*).

Experiment Design

A 3 (Takeover Scenarios) by 7 (NDRTs) within-subject design was used. Each combination of NDRT type and scenario type happened in one drive, leading to a total of 21 drives. Each drive was around 7 minutes long, and the takeover scenarios happened near the end of the drive. The 3 takeover scenarios were counterbalanced using a Latin squared design. For each kind of takeover scenario, the 7 NDRT types were also counterbalanced using a Latin squared design, leading to 21 unique experimental orders. Each order was experienced by two participants. Further, after completing all 7 drives for one kind of takeover scenario, participants took a 10-minute break before proceeding to the next 7 drives.

Procedure

Participants followed pre-experiment instructions and completed a 30-minute orientation. Then, physiological sensors (ECG, RESP, EDA) and eye-tracking devices were put on and calibrated before the formal drives. All the physiological data was collected at 100 Hz.

Signal Preprocessing

Given that we used the raw data as inputs in the model, only noise elimination was conducted to enhance the quality of the data. Specifically, all signals underwent down-sampling to a frequency of 100 Hz (from 1000Hz in MADT-D) to optimize computational efficiency and ensure the consistency between two datasets. For EDA, a low-pass filter with a cutoff frequency of 5 Hz was employed; while for ECG and RESP, band-pass filters were applied within the frequency ranges of 3 Hz to 45 Hz and 0.1 Hz to 0.35 Hz, respectively (9). The preprocessing was executed using Python 3.8.

RESULTS

Experimental Results and Analysis

We compared our model to selected learning-based approaches from prior studies for cognitive load estimation in driving, including MTS-CNN (*19*), DecNet (*20*), CNN-LSTM (*21*), m-HyperLSTM (*22*) and ARecNet (*23*). As shown in Table 1, the model comparison encompassed various cognitive load tasks, and classification categories, showcasing the model performance across different time horizons (i.e., the length of physiological signal time series, represented by t_w).

It can be found that m-HyperLSTM consistently outperformed MTS-CNN and DecNet. Nevertheless, our proposed CogFormer consistently achieved the highest accuracy in all models. Furthermore, we observed that for some tasks, increasing the time horizon did not necessarily lead to an increase in recognition accuracy. The increased time horizon may have introduced some noise in the inputs for some tasks. Although the CogFormer can capture long-term temporal information, it may still not be capable enough to automatically pay attention to high-value information in the data. An improved attention mechanism may be needed to improve the capability of the model in filtering noise information in long-term temporal sequences.

TABLE 1 Comparison within-subject cognitive load estimation.

Model	$t_w = 1\ s$			$t_w = 3\ s$			$t_w = 5\ s$		
	Accuracy (%)	F1-score	AUC	Accuracy (%)	F1-score	AUC	Accuracy (%)	F1-scores	AUC
MADT-D: Math task (two classes)									
MTS-CNN$^{\Delta}$	87.71 ± 4.86	0.876 ± 0.044	0.877 ± 0.046	88.83 ± 4.14	0.884 ± 0.042	0.882 ± 0.049	88.29 ± 4.03	0.887 ± 0.042	0.885 ± 0.046
DecNet$^{\Delta}$	86.76 ± 3.29	0.866 ± 0.035	0.863 ± 0.033	87.27 ± 4.47	0.877 ± 0.046	0.876 ± 0.041	87.45 ± 4.34	0.876 ± 0.043	0.872 ± 0.049
CNN-LSTM$^{\Delta}$	87.51 ± 4.14	0.871 ± 0.043	0.879 ± 0.042	87.87 ± 5.63	0.876 ± 0.053	0.874 ± 0.048	88.38 ± 3.59	0.888 ± 0.033	0.881 ± 0.036
m-HyperLSTM$^{\Delta}$	89.68 ± 5.26	0.898 ± 0.055	0.896 ± 0.053	89.32 ± 5.19	0.894 ± 0.053	0.881 ± 0.051	89.14 ± 4.72	0.892 ± 0.045	0.893 ± 0.046
ARecNet$^{\Phi}$	92.46 ± 4.22	0.921 ± 0.045	0.919 ± 0.040	91.93 ± 3.12	0.913 ± 0.031	0.911 ± 0.033	92.56 ± 3.38	0.921 ± 0.031	0.922 ± 0.032
CogFormer (ours)$^{\Phi}$	**95.28** ± 3.98	**0.952** ± 0.039	**0.955** ± 0.037	**93.79** ± 4.25	**0.938** ± 0.047	**0.936** ± 0.044	**94.03** ± 3.81	**0.944** ± 0.039	**0.947** ± 0.035
CAM-CLD: Spacial task (two classes)									
MTS-CNN$^{\Delta}$	86.08 ± 2.85	0.863 ± 0.026	0.865 ± 0.027	87.34 ± 2.39	0.874 ± 0.022	0.879 ± 0.021	88.57 ± 1.97	0.882 ± 0.022	0.885 ± 0.021
DecNet$^{\Delta}$	86.91 ± 2.43	0.867 ± 0.028	0.863 ± 0.023	86.23 ± 1.47	0.867 ± 0.017	0.866 ± 0.016	88.06 ± 2.44	0.888 ± 0.024	0.882 ± 0.025
CNN-LSTM$^{\Delta}$	89.46 ± 2.32	0.899 ± 0.027	0.898 ± 0.026	88.41 ± 2.52	0.887 ± 0.027	0.884 ± 0.023	89.46 ± 2.37	0.892 ± 0.026	0.894 ± 0.023
m-HyperLSTM$^{\Delta}$	88.64 ± 3.76	0.884 ± 0.034	0.889 ± 0.036	88.96 ± 1.29	0.886 ± 0.016	0.884 ± 0.014	89.32 ± 3.45	0.898 ± 0.032	0.893 ± 0.036
ARecNet$^{\Phi}$	91.92 ± 2.46	0.912 ± 0.027	0.909 ± 0.028	90.09 ± 1.33	0.897 ± 0.015	0.890 ± 0.013	91.49 ± 2.96	0.916 ± 0.029	0.918 ± 0.027
CogFormer (ours)$^{\Phi}$	**93.05** ± 2.39	**0.932** ± 0.026	**0.931** ± 0.023	**91.68** ± 1.68	**0.918** ± 0.013	**0.921** ± 0.018	**93.63** ± 2.12	**0.934** ± 0.023	**0.931** ± 0.022
CAM-CLD: Math task (three classes)									
MTS-CNN$^{\Delta}$	85.16 ± 3.18	0.854 ± 0.035	0.852 ± 0.036	86.28 ± 3.13	0.867 ± 0.034	0.864 ± 0.032	86.99 ± 2.35	0.868 ± 0.021	0.865 ± 0.026
DecNet$^{\Delta}$	85.59 ± 2.75	0.858 ± 0.025	0.853 ± 0.026	86.64 ± 2.92	0.863 ± 0.027	0.863 ± 0.029	87.37 ± 2.51	0.870 ± 0.025	0.872 ± 0.024
CNN-LSTM$^{\Delta}$	86.86 ± 2.27	0.864 ± 0.022	0.862 ± 0.021	87.59 ± 2.23	0.879 ± 0.021	0.877 ± 0.023	87.98 ± 3.12	0.879 ± 0.034	0.872 ± 0.037
m-HyperLSTM$^{\Delta}$	87.99 ± 2.36	0.872 ± 0.025	0.876 ± 0.023	87.87 ± 2.13	0.875 ± 0.023	0.879 ± 0.019	88.25 ± 2.55	0.888 ± 0.022	0.885 ± 0.027
ARecNet$^{\Phi}$	88.28 ± 2.16	0.889 ± 0.020	0.884 ± 0.022	88.17 ± 3.37	0.883 ± 0.039	0.882 ± 0.032	89.23 ± 2.13	0.894 ± 0.025	0.901 ± 0.022
CogFormer (ours)$^{\Phi}$	**91.13** ± 1.31	**0.909** ± 0.012	**0.910** ± 0.015	**90.64** ± 2.43	**0.902** ± 0.023	**0.904** ± 0.022	**91.92** ± 2.36	**0.920** ± 0.022	**0.921** ± 0.026
CAM-CLD: n-back task (four classes)									
MTS-CNN$^{\Delta}$	84.35 ± 3.08	0.841 ± 0.032	0.842 ± 0.031	84.93 ± 2.55	0.846 ± 0.026	0.845 ± 0.023	85.33 ± 2.14	0.856 ± 0.023	0.852 ± 0.021
DecNet$^{\Delta}$	84.26 ± 2.09	0.848 ± 0.021	0.845 ± 0.019	85.01 ± 2.58	0.855 ± 0.025	0.851 ± 0.026	85.43 ± 2.33	0.851 ± 0.024	0.856 ± 0.024
CNN-LSTM$^{\Delta}$	85.88 ± 2.94	0.859 ± 0.029	0.850 ± 0.030	85.12 ± 2.64	0.858 ± 0.024	0.852 ± 0.025	85.36 ± 2.19	0.858 ± 0.023	0.851 ± 0.021
m-HyperLSTM$^{\Delta}$	85.85 ± 2.57	0.854 ± 0.025	0.857 ± 0.026	84.47 ± 2.26	0.845 ± 0.023	0.841 ± 0.021	86.17 ± 2.21	0.864 ± 0.024	0.867 ± 0.026
ARecNet$^{\Phi}$	86.17 ± 1.52	0.862 ± 0.016	0.869 ± 0.014	85.13 ± 1.54	0.858 ± 0.016	0.852 ± 0.018	87.42 ± 1.74	0.872 ± 0.017	0.870 ± 0.016
CogFormer (ours)$^{\Phi}$	**88.72** ± 1.78	**0.890** ± 0.019	**0.892** ± 0.018	**87.67** ± 1.44	**0.878** ± 0.014	**0.872** ± 0.015	**88.14** ± 1.88	**0.882** ± 0.018	**0.885** ± 0.017

Notes: Δ means feature-level fusion model, Φ means decision-level fusion model.

Ablation Experiment

Ablation experiments revealed that the Aligned Attention module consistently enhances spatial feature capture, outperforming traditional transformers. Multi-stream Encoding excels with shorter historical horizons, while concatenated encoding improves with longer horizons. Results indicate that Aligned Attention strengthens multimodal information representation in extended sequences (see Table 2).

TABLE 2 Ablation study with different historical horizons on recognition accuracy with various cognitive tasks.

	Variants		CogFormer
Multi-stream Encoding	×	√	√
Aligned Attention	×	×	√
MADT-D: Math task			
$t_w = 1\ s$	91.41 (↓3.87%)	91.85 (↓**3.43%**)	**95.28**
$t_w = 3\ s$	90.47 (↓**3.32%**)	90.64 (↓3.55%)	**93.79**
$t_w = 5\ s$	91.58 (↓**2.45%**)	91.06 (↓2.97%)	**94.03**
CAM-CLD: Spacial task			
$t_w = 1\ s$	88.51 (↓4.54%)	89.23 (↓**3.82%**)	**93.05**
$t_w = 3\ s$	88.03 (↓3.65**%**)	88.34 (↓**3.34%**)	**91.68**
$t_w = 5\ s$	90.61 (↓**3.02%**)	90.50 (↓3.13%)	**93.63**
CAM-CLD: Math task			
$t_w = 1\ s$	87.70 (↓3.43%)	87.87 (↓**3.26%**)	**91.13**
$t_w = 3\ s$	85.99 (↓4.65%)	86.99 (↓**3.65%**)	**90.64**
$t_w = 5\ s$	88.10 (↓3.82%)	88.21 (↓**3.71%**)	**91.92**
CAM-CLD: n-back task			
$t_w = 1\ s$	82.96 (↓5.76%)	84.18 (↓**4.54%**)	**88.72**
$t_w = 3\ s$	82.96 (↓4.71%)	83.66 (↓**4.01%**)	**87.67**
$t_w = 5\ s$	84.91 (↓**3.23%**)	84.93 (↓3.59%)	**88.14**

Robustness Testing

In practical applications, time series signals are affected by missing values and Gaussian White Noise (GWN). Robustness tests showed that CogFormer consistently outperforms ARecNet in handling these distortions, except slightly in the math task under mixed conditions. This highlights CogFormer's superior robustness with its parallel transformer and coherent attention mechanism, see Table 3.

TABLE 3 Comparison of recognition accuracy in the decision-fusion models with varied information distortions when tw=5s.

Model	Normal (%)	Missing 10% ($\mathcal{M}$)	Missing 20% ($\mathcal{M}$)	Missing 30% (α)	Missing 30% (β)	Missing 30% (γ)	GWN σ=0.1	GWN σ=0.2	Mixed distortion
MADT-D: Math task									
CogFormer (Proposed)	94.03	92.80 (↓1.23)	91.36 (↓2.67)	92.39 (↓1.64)	92.70 (↓1.33)	92.74 (↓1.29)	93.61 (↓0.42)	92.30 (↓1.73)	90.80 (↓**3.23**)
ARecNet (Contrastive)	92.56	90.81 (↓1.75)	88.75 (↓3.81)	90.20 (↓2.36)	90.35 (↓2.21)	91.18 (↓1.38)	91.85 (↓0.71)	90.00 (↓2.56)	87.14 (↓5.42)
CAM-CLD: Spacial task									
CogFormer (Proposed)	93.63	91.87 (↓1.76)	91.51 (↓2.12)	91.86 (↓1.77)	92.39 (↓1.24)	91.60 (↓2.03)	93.27 (↓0.36)	91.18 (↓2.45)	89.61 (↓**4.02**)
ARecNet (Contrastive)	91.49	89.36 (↓2.13)	87.98 (↓3.51)	89.52 (↓1.97)	89.46 (↓1.03)	88.94 (↓2.55)	90.82 (↓0.67)	88.33 (↓3.16)	86.87 (↓4.62)
CAM-CLD: Math task									
CogFormer (Proposed)	91.92	90.29 (↓1.63)	88.81 (↓3.11)	90.18 (↓1.74)	90.79 (↓1.13)	89.77 (↓2.15)	91.29 (↓0.63)	89.76 (↓2.16)	86.39 (↓5.53)
ARecNet (Contrastive)	89.23	87.71 (↓1.52)	86.21 (↓3.02)	86.78 (↓1.45)	87.31 (↓1.92)	86.77 (↓2.46)	88.88 (↓0.35)	86.47 (↓1.76)	83.99 (↓**5.24**)
CAM-CLD: n-back task									
CogFormer (Proposed)	88.14	85.50 (↓2.64)	83.71 (↓4.43)	85.83 (↓2.31)	86.12 (↓2.02)	86.00 (↓2.14)	87.22 (↓0.92)	84.80 (↓3.34)	81.73 (↓**6.41**)
ARecNet (Contrastive)	87.42	83.89 (↓3.53)	82.67 (↓4.75)	85.25 (↓2.17)	85.89 (↓1.53)	85.22 (↓2.20)	86.09 (↓1.33)	83.41 (↓4.01)	80.20 (↓7.22)

Note: ECG, RESP, EDA - $\mathcal{M}$, ECG -α, RESP -β, and EDA -γ, GWN -σ.

DISCUSSION AND CONCLUSION

Being different from previous approaches that utilized traditional machine/deep learning models, we developed a decision-level multi-physiological information fusion architecture to extract temporal and spatial information from multiple physiological signals. Experimental results demonstrate that the proposed CogFormer surpassed other baseline models in terms of estimation accuracy and robustness. In addition, based on the model ablation study and robustness test, we find that:

The preferred feature combinations for driver state estimation may depend on the type of targeted tasks, data collection quality, and driving context. Thus, the performance of models based on handcrafted features and manual feature selection may not be guaranteed in real-world applications.

- A longer time horizon, though can provide richer information, may not necessarily increase the model performance, potentially because the models may not be able to capture the complex features in the data. Thus, the degree of matching between the models and the characteristics of data should be considered when designing driver state-monitoring algorithms.

- All models, including our proposed model, were highly susceptible to individual differences - the performance of all models dropped significantly when across-subjects data partition was applied, indicating that the models, even with the attention mechanism, may still not be able to capture the individual-invariant features of high cognitive load states. Future research should design specific algorithms and structures to handle this issue.
- Though our algorithm was not designed specifically to handle the noise and data distortion, our model showed better robustness compared to the best baseline model. The noise and missing data are common in real-world applications. Thus, future research should consider a more specific algorithm design and validate the proposed model based on real-world datasets.

REFERENCES

1. Singh, S. Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey. *Traffic Safety Facts - Crash Stats*, 2015.
2. Jackson, L., P. Chapman, and D. Crundall. What Happens next? Predicting Other Road Users' Behaviour as a Function of Driving Experience and Processing Time. *Ergonomics*, Vol. 52, No. 2, 2009, pp. 154–164. https://doi.org/10.1080/00140130802030714.
3. Harbluk, J. L., Y. I. Noy, P. L. Trbovich, and M. Eizenman. An On-Road Assessment of Cognitive Distraction: Impacts on Drivers' Visual Behavior and Braking Performance. *Accident Analysis & Prevention*, Vol. 39, No. 2, 2007, pp. 372–379. https://doi.org/10.1016/j.aap.2006.08.013.
4. Recarte, M. A., and L. M. Nunes. Effects of Verbal and Spatial-Imagery Tasks on Eye Fixations While Driving. *Journal of Experimental Psychology: Applied*, Vol. 6, No. 1, 2000, pp. 31–43. https://doi.org/10.1037/1076-898X.6.1.31.
5. Muhrer, E., and M. Vollrath. The Effect of Visual and Cognitive Distraction on Driver's Anticipation in a Simulated Car Following Scenario. *Transportation Research Part F: Psychology and Behaviour*, Vol. 14, No. 6, 2011, pp. 555–566. https://doi.org/10.1016/j.trf.2011.06.003.
6. Huang, Chunxi, Xie, Weiyin, Huang, Qihao, Zhu, Yan, Cui, Dixiao, and He, Dengbo. Effect of Advanced Driver Assistance Systems on Fatigue Levels of Heavy Truck Drivers in Prolonged Driving Tasks. *Journal of Tongji University (Natural Science)*, Vol. 52, No. 6, 2024, pp. 846–855.
7. Stapel, J., F. A. Mullakkal-Babu, and R. Happee. Automated Driving Reduces Perceived Workload, but Monitoring Causes Higher Cognitive Load than Manual Driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 60, 2019, pp. 590–605. https://doi.org/10.1016/j.trf.2018.11.006.
8. Wiediartini, U. Ciptomulyono, and R. S. Dewi. Evaluation of Physiological Responses to Mental Workload in N-Back and Arithmetic Tasks. *Ergonomics*, 2023, pp. 1–13. https://doi.org/10.1080/00140139.2023.2284677.
9. Meteier, Q., M. Capallera, S. Ruffieux, L. Angelini, O. Abou Khaled, E. Mugellini, M. Widmer, and A. Sonderegger. Classification of Drivers' Workload Using Physiological Signals in Conditional Automation. *Frontiers in Psychology*, Vol. 12, 2021, p. 596038. https://doi.org/10.3389/fpsyg.2021.596038.
10. Shi, W., Z. Wang, A. Wang, and D. He. Classification of Driver Cognitive Load in Conditionally Automated Driving: Utilizing Electrocardiogram-Based Spectrogram with

Lightweight Neural Network. *Transportation Research Record*, 2024. https://doi.org/10.1177/03611981241252797.
11. Wang, A., J. Wang, W. Shi, and D. He. Cognitive Workload Estimation in Conditionally Automated Vehicles Using Transformer Networks Based on Physiological Signals. *Transportation Research Record*, 2024. https://doi.org/10.1177/03611981241250023.
12. Kumar, S., D. He, G. Qiao, and B. Donmez. Classification of Driver Cognitive Load Based on Physiological Data: Exploring Recurrent Neural Networks. Presented at the 2022 International Conference on Advanced Robotics and Mechatronics (ICARM), Guilin, China, 2022.
13. He, D., Z. Wang, E. B. Khalil, B. Donmez, G. Qiao, and S. Kumar. Classification of Driver Cognitive Load: Exploring the Benefits of Fusing Eye-Tracking and Physiological Measures. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2676, No. 10, 2022, pp. 670–681. https://doi.org/10.1177/03611981221090937.
14. Qu, Y., H. Hu, J. Liu, Z. Zhang, Y. Li, and X. Ge. Driver State Monitoring Technology for Conditionally Automated Vehicles: Review and Future Prospects. *IEEE Transactions on Instrumentation and Measurement*, Vol. 72, 2023, pp. 1–20. https://doi.org/10.1109/TIM.2023.3301060.
15. Ansari, S., F. Naghdy, H. Du, and Y. Pahnwar. Driver Mental Fatigue Detection Based on Head Posture Using New Modified reLU-BiLSTM Deep Neural Network. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, Vol. 23, No. 8, 2022, pp. 10957–10969. https://doi.org/10.1109/TITS.2021.3098309.
16. Meteier, Q., M. Capallera, E. de Salis, L. Angelini, S. Carrino, M. Widmer, O. Abou Khaled, E. Mugellini, and A. Sonderegger. A Dataset on the Physiological State and Behavior of Drivers in Conditionally Automated Driving. *Data in Brief*, Vol. 47, 2023. https://doi.org/10.1016/j.dib.2023.109027.
17. Jaeggi, S. M., M. Buschkuehl, W. J. Perrig, and B. Meier. The Concurrent Validity of the N-Back Task as a Working Memory Measure. *Memory*, Vol. 18, No. 4, 2010, pp. 394–412. https://doi.org/10.1080/09658211003702171.
18. Liang, Y., and J. D. Lee. Combining Cognitive and Visual Distraction: Less than the Sum of Its Parts. *Accident Analysis & Prevention*, Vol. 42, No. 3, 2010, pp. 881–890. https://doi.org/10.1016/j.aap.2009.05.001.
19. Xie, Y., Y. L. Murphey, and D. S. Kochhar. Personalized Driver Workload Estimation Using Deep Neural Network Learning From Physiological and Vehicle Signals. *IEEE Transactions on Intelligent Vehicles*, Vol. 5, No. 3, 2020, pp. 439–448. https://doi.org/10.1109/TIV.2019.2960946.
20. Amadori, P. V., T. Fischer, R. Wang, and Y. Demiris. Predicting Secondary Task Performance: A Directly Actionable Metric for Cognitive Overload Detection. *IEEE Transactions on Cognitive and Developmental Systems*, Vol. 14, No. 4, 2022, pp. 1474–1485. https://doi.org/10.1109/TCDS.2021.3114162.
21. Huang, J., Y. Liu, and X. Peng. Recognition of Driver's Mental Workload Based on Physiological Signals, a Comparative Study. *BIOMEDICAL SIGNAL PROCESSING AND CONTROL*, Vol. 71, 2022. https://doi.org/10.1016/j.bspc.2021.103094.
22. Wang, R., P. V. Amadori, and Y. Demiris. Real-Time Workload Classification during Driving Using HyperNetworks. Presented at the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018.

23. Yang, H., J. Wu, Z. Hu, and C. Lv. Real-Time Driver Cognitive Workload Recognition: Attention-Enabled Learning with Multimodal Information Fusion. *IEEE Transactions on Industrial Electronics*, 2023, pp. 1–11. https://doi.org/10.1109/TIE.2023.3288182.