Dear Editor,

We thank you and the reviewers for the valuable comments and suggestions on our manuscript titled “Multimodal Fusion of Heart Rate and Gaze Data for Real-Time Driver Monitoring in Naturalistic Driving.” We have carefully revised our manuscript according to each reviewer’s feedback. Below is our detailed response addressing each point raised.

Reviewer 1

Comment 1:

“The research uses a heart rate strap to measure the driver’s heart rate data. However, the sampling rate of the strap is only 1 Hz, which is insufficient for further heart rate analysis.”

Response 1:

We clearly mentioned in Section 3.2 that the Polar H10 sensor records raw ECG data at 1000 Hz, as supported by peer-reviewed literature [1]. The 1 Hz data reported in our analysis represent processed heart rate (HR) values, which are sufficient to monitor driver HR changes. The confusion might come from the system model (Fig.1), which is now updated and removed (1Hz). Additionally, we have updated Section 3.2 to provide further clarification of our perspective.

> *“The Polar H10 records raw electrocardiogram (ECG) signals at a sampling frequency of 1000 Hz [36]. In this study, HR data were processed to output an HR value at 1 Hz, which is sufficient for capturing temporal changes in HR during real-world driving.”*

Comment2:

“The number of participants is only five, which is too small to ensure the generalizability of the experimental results. Additionally, the detailed information of the participants (e.g., age, gender, years of driving experience) is not provided.”

Response 2:

This study is indeed a pilot, yet it includes a substantial amount of data, totaling 480 minutes. The extensive recording provides sufficient data points for preliminary validation. Section 3.1 has been updated to provide a clearer explanation of our work.

> *“Although the presented work here is a pilot study, it includes a substantial amount of driving data, totaling 480 minutes, which provides sufficient data points for preliminary validation and statistical analysis.”*

A table detailing participant demographics has been added to the manuscript (as Table 2):

Participant	***Age***	***Gender***	***Driving Experience (Years)***
1	*35-40*	*Male*	*15-20*
2	*25-30*	*Male*	*5-10*
3	*30-35*	*Male*	*5-10*
4	*30-35*	*Male*	*10-15*
5	*40-45*	*Male*	*20-25*

Comment 3:

“Although the heart rate strap and the glasses-type eye tracker have less impact on participants than other sensors, they still introduce certain disturbances. For example, drivers who are not accustomed to wearing glasses may experience discomfort on the bridge of the nose after prolonged use (especially when an additional pair of glasses is required for those who wear glasses). The heart rate strap may interfere with vehicle operation due to skin contact. Therefore, a familiarization period should be included before the experiment begins to allow drivers to adapt to the equipment. This step is missing in the paper and is recommended for future studies.”

Response 3:

Participants underwent a familiarization period in a parking lot, wearing all sensors for at least 20 minutes while the copilot confirmed comfort and proper sensor function. Participants also drove within the parking area and only proceeded when they felt safe and comfortable. These details have also been added in Section 3.1.

> *“Before commencing the actual drives, participants underwent a familiarization period in a parking lot. During this period, all sensors were worn for at least 20 minutes while the copilot ensured correct sensor operation and participant comfort. Participants also drove short loops within the parking area to acclimate to the equipment, proceeding to public roads only when they felt safe and comfortable.”*

Comment 4:

“When using physiological data to describe a driver's state, heart rate (HR) varies greatly among individuals, making the evaluation results unreliable. Heart rate variability (HRV) should be used instead during the evaluation process.”

Response 4:

This study focuses specifically on HR. HRV can offer additional insights and will be considered in future studies. Also, state-of-the-art research [2] supports the higher accuracy of HR compared to HRV in detecting cognitive load during urban and motorway driving scenarios. [Supporting statement is added in Section 3.2]

> *"This study focuses on HR rather than heart rate variability (HRV). While HRV can provide additional information on autonomic nervous system activity, our focus is motivated by the fact that HR alone has demonstrated strong sensitivity to cognitive and environmental driving demands. State-of-the-art research [37] has shown that HR can achieve higher accuracy than HRV in differentiating cognitive load during urban and motorway driving scenarios."*

Comment 5:

"What were the weather conditions during the experiment? Please also list the distribution of these conditions."

Response 5:

Data collection took place during the daytime under various conditions: clear, cloudy, and light rain. Extreme conditions (such as heavy rain and fog) were not included. Section 3.1 has been updated to clearly declare the environment.

> *"Data collection took place mostly during the daytime and under various weather conditions, including clear skies, overcast conditions, and light rain. Extreme weather scenarios (e.g., heavy rain, fog) were intentionally excluded to maintain consistent visual and physiological measurement conditions."*

Comment 6:

"Please explain the rationale for dividing the study period into 10-second intervals. Also, explain how the key metrics were selected and whether the selection is reasonable."

Response 6:

Although the proposed algorithm can adapt to any defined window, a 10-second window effectively detects real-time HR changes, matching the Polar sensor's internal processing and the real-time metrics from the gaze glasses. Subsection 3.5 has been updated. Longer windows may capture more context, but they would reduce the system's sensitivity [3]. Optimal window size adjustment for each modality will be considered in future work. We updated and mentioned such a limitation in the conclusion.

> *"Although the proposed algorithm can be adapted to any window length, a 10-second duration has been shown to effectively capture short-term changes in physiological and visual attention, while remaining responsive to transient events. Longer windows could capture more context but would reduce the temporal sensitivity of the system [42]"*

Comment 7:

“Please explain how the parameter θ in the algorithm was determined, and how the weights of the various coefficients were decided.”

Response 7:

The parameters and weights in this pilot study were initially determined through statistical analyses. [Justification in Section 3.5 and analyses of different values are in Section 4.2.3.] This initial selection lays the groundwork for future work to refine these parameters through detailed scenario labeling and analysis.

> *“The metric weights were initially determined based on statistical analyses of our dataset (Section 4.1), with additional validation through sensitivity analysis (Section 4.2.3).”*

Comment 8:

“The proposed algorithm uses 0-1 variables to determine whether the sum of several weighted parameters exceeds a certain threshold as a judgment criterion. This method is not reasonable.”

Response 8:

A binary classification is selected intentionally for real-time practicality, facilitating rapid response to safety-critical situations instead of multiple class classifications. Recent literature [4] confirms the effectiveness of threshold-based anomaly detection methods (binary, distinguishing between normal and abnormal) in mental state monitoring. This statement and reference are added in Section 3.5.

> *“A binary classification (Normal vs. Abnormal) was intentionally adopted for real-time practicality, as it facilitates rapid intervention in safety-critical contexts without the computational overhead of multi-class classification [46].”*

Reviewer 2:

Comment 1:

“This work seems just collected 5 drivers data, and using the collected data to do some simple analysis, and got some common conclusion. For example, fig.3 and fig.5,fig.8 and fig.9, after the data collected, just very simple processing can get these results. And the collusion for each results also not novel.”

Response 1:

The novelty of our work lies in our effort to integrate physiological and gaze-tracking metrics to monitor driver health and behavior comprehensively. Moreover, we proposed a weighting mechanism that lays the groundwork for future refinements to link which metric is more correlated with the driver state.

Comment 2:

“The title of this paper is multimodal fusion of XXXX , where is the contribution for the “fusion”? the HR data and gaze data both were analyzed separately. May be the author think the table 4 is the “fusion” part, but it was not enough.”

Response 2:

The fusion occurs at the decision level through a weighted algorithm that integrates HR and gaze metrics, laying the groundwork for future refinements in parameter tuning and enhancing driver status detection by using refined parameters for each modality based on labeled scenarios and deep learning outcomes. The decision fusion is explained in line 230.

> *“The fusion occurs through a weighted decision-making process, combining HR and gaze metrics into a unified state score. This approach not only enables real-time driver state estimation but also establishes a foundation for future refinements in parameter tuning based on scenario-specific labeling and deep learning models.”*

Comment 3:

“the experiment for 5 drivers data detection is done under the same traffic flow? High traffic flow and light traffic flow may have quite different effects for drivers.”

Response 3:

These factors are indeed limitations. Future extensive data collection will include detailed traffic density analysis and controlled scenarios to improve generalizability. However, information about the environment during data collection is included in Section 3.1, and the table for participants' details has been updated to include a driving experience column.

> *“Traffic conditions were naturally variable, ranging from low-density rural motorway segments to dense urban traffic, and included both high-flow and stop-and-go conditions.”*

Comment 4:

“The 5 drivers have different driving experiences, this is also have the effects on the collected data, such as in line 346, subject 3 has less driving experience and subject 5 have extensive experiences. This will lead to the collected data is not a common results and the results from these data are not unreliable.”

Response 4:

Driving experience is included in the participants' table, which is added to the manuscript. This is a pilot study and initial setup; however, such a variation among participants will be considered for our final data collection protocol.

Comment 5:

“5 subjects are also a very small sample, can not enough to reveal some reliable conclusions. More subjects should be considered.”

Response 5:

This study is indeed a pilot, yet it includes a substantial amount of data, totaling 480 minutes. The extensive recording provides sufficient data points for preliminary validation. Section 3.1 has been updated to give a more precise explanation of our work. Also, a table detailing participant demographics has been added to the manuscript (as Table 2):

Reviewer 3:

Comment 1:

“The duration of the driving is long, so despite the small number of participants, a comparison can be made between the motorway scenario and the urban scenario. However, there is no precise definition of the criteria for “normal“ and “abnormal“, and there is no ground truth for this segmentation. The authors said: “To ensure labeling consistency, approximately 5% of segments were manually reviewed by the first author with the assistance of dashcam recordings captured during data collection.” It is not clear according to which criteria the manual verification was performed. Also, if the whole signal was only automatically labeled (except 5% of the signal), how can we be sure that the labeling was right?”

Response 1:

Manual verification during this pilot involved examining the environmental context (traffic density and stops at traffic lights). Normal and abnormal are defined in Algorithm1. More accurate manual labeling will be conducted in future studies for comprehensive validation. A statement of such limitations is revised in the conclusion.

> *“Despite these positive findings, several limitations need to be investigated further. First, as this is a pilot study, it involved a small number of participants; a larger pool of volunteers will be recruited as the project progresses. Second, the manually chosen thresholds and metric weights, although based on statistical calculations and literature, might not be fully reflective of variations in individual physiology or driving behavior. Third, although naturalistic driving data improves ecological validity, it introduces variability that demands robust experimental design and well-defined labeling procedures to ensure consistency and reliability.”*

Comment 2:

“The study includes human participants, but the ethical approval and informed consent were not mentioned in the text. Also, the participant description is not provided (what was the gender and age distribution of the participants). The driver’s experience should be provided for all participants, not only for some of them.”

Response 2:

Participants include some of the authors and friends, who provided informal consent, understanding the voluntary nature of their participation. Formal ethical approval was not obtained at this preliminary stage; future studies with broader data collection will ensure formal ethical compliance. Section 3.1 is updated.

> *“All participants provided informal consent before the study, acknowledging the voluntary nature of participation and their understanding of the procedures involved. Given the pilot nature of this work and the use of commercially available sensors, formal institutional ethical approval was not sought at this stage; however, future studies involving broader participant recruitment will follow formal ethical approval protocols.”*

Comment 3:

“Details about the distribution of motorway/urban driving are not provided. How was the driving organized? Did participants drive first in one scenario and then in another? Or were these scenarios mixed? How many times does one participant drive in a motorway/urban scenario during the total time of 480 min? This distribution of scenarios can have a significant impact on the results. Was the participant alone in the car? How did the weather change during the drive? How did the traffic change during the drive? All these factors should be included in the manual labeling of segments.”

Response 3:

Driving scenarios (city and motorway) were naturally mixed, with some repeated on different days, reflecting realistic conditions. The dataset comprises approximately 33% motorway and 66% city driving. Additionally, as mentioned in Section 3.1, the copilot is responsible for data collection, ensuring that each participant is accompanied and not alone in the car. Section 3.1 is updated.

> *“Driving scenarios were naturally mixed, with participants alternating between city and motorway environments over multiple sessions and on different days. The final dataset comprises approximately 66% urban driving and 33% motorway driving.”*

Comment 3:

“The references for the lower and upper bounds are given for blink duration, fixation duration, and saccade amplitude, but for other features it is not clear how the bounds are adopted. Without the baseline recording of the HR and normalization of HR to the baseline, it would not even be possible to set such bounds (due to the natural HR variability between participants).”

Response 3:

Thresholds for HR and saccade velocity and duration were explicitly derived from the distribution analysis presented in Section 4.1 and justified in lines 346-348.

> *“...segments with mean HR between 65–95 bpm, fixation duration between 150–900 ms, saccade duration between 0–100 ms, saccade amplitude between 0°–*

> *15°, and saccade velocity between 0–3000 px/s were defined as normal (see Table 4).”*

Comment 4:

“It is not clear how the distribution of weight factors is selected. How did you prove that the selected ratio of weights is right?”

Response 4:

The parameters and weights in this pilot study were initially determined through statistical analyses. [Justification in Section 3.5 and analyses of different values are in Section 4.2.3.] This initial selection lays the groundwork for future work to refine these parameters through detailed scenario labeling and analysis.

> *“The metric weights were initially determined based on statistical analyses of our dataset (Section 4.1), with additional validation through sensitivity analysis (Section 4.2.3).”*

Comment 5:

“There is no discussion according to the state-of-the-art literature.”

Response 5:

Our manuscript includes 50 references, including papers from this year, 2025. The related work section provides a comprehensive overview of the current state of driver monitoring, accompanied by a table comparing our work to the existing state of the art.

We sincerely appreciate the reviewers' insightful suggestions, which have helped us improve our manuscript. Should further clarification or additional information be required, please do not hesitate to contact us.

Thank you for considering our revised manuscript for publication.

References

[1] Schaffarczyk, M.; Rogers, B.; Reer, R.; Gronwald, T. Validity of the polar H10 sensor for heart rate variability analysis during resting state and incremental exercise in recreational men and women. Sensors 2022, 22, 6536.

[2] Arutyunova, K.R., Bakhchina, A.V., Konovalov, D.I. *et al.* Heart rate dynamics for cognitive load estimation in a driving simulation task. *Sci Rep* **14**, 31656 (2024).

[3] S. Chakraborty, P. Kiefer and M. Raubal, "Estimating Perceived Mental Workload From Eye-Tracking Data Based on Benign Anisocoria," in *IEEE Transactions on Human-Machine Systems*, vol. 54, no. 5, pp. 499-507, Oct. 2024.

[4] Visconti, P.; Rausa, G.; Del-Valle-Soto, C.; Velázquez, R.; Cafagna, D.; De Fazio, R. Innovative Driver Monitoring Systems and On-Board-Vehicle Devices in a Smart-Road Scenario Based on the Internet of Vehicle Paradigm: A Literature and Commercial Solutions Overview. *Sensors* **2025**, *25*, 562.