Gaze-based Drowsiness Recognition for Level 3 Automated Driving: Designing Core Features and Temporal Windows

Eun Hye Jang
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
cleta4u@etri.re.kr

Mi Chang
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
rosechang@etri.re.kr

Woojin Kim
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
wjinkim@etri.re.kr

Jiwoo Han
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
chau@etri.re.kr

Daesub Yoon
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
eyetracker@etri.re.kr

Abstract

In Level 3 automated driving, drivers must supervise the system and resume control when requested, making drowsiness a critical safety risk. This study investigates gaze-based drowsiness recognition, focusing on selecting core gaze features and configuring temporal windows for real-time monitoring. Eye-tracking data were collected from 100 participants while they experienced automated driving in a simulator equipped with production vehicle components. Drowsiness was assessed using Karolinska Sleepiness Scale ratings every five minutes. Gaze-based features were extracted over multiple windows and refined through correlation analysis, multicollinearity reduction, model-based feature importance, and cross-window reliability analysis. Four core features-pupil variability, mean fixation duration, eyelid openness, and vertical gaze dispersion-were identified, with a 60-second window providing the best reliability and 30-second windows remaining viable when shorter detection latency is required. A Random Forest classifier achieved an AUC of 0.832 using the four features over 60 seconds. These findings offer design-oriented guidance for implementing gaze-based drowsiness recognition in Level 3 automated driving systems.

CCS Concepts

• **Human-centered computing** → Human-computer interaction (HCI); Empirical studies in HCI.

Keywords

Driver drowsiness recognition, Gaze-based monitoring, Supervisory driving, Temporal window selection

ACM Reference Format:
Eun Hye Jang, Mi Chang, Woojin Kim, Jiwoo Han, and Daesub Yoon. 2026. Gaze-based Drowsiness Recognition for Level 3 Automated Driving: Designing Core Features and Temporal Windows. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3772363.3798868

ACM ISBN 979-8-4007-2281-3/2026/04
https://doi.org/10.1145/3772363.3798868

1 Introduction

In SAE Level 3 automated driving, drivers supervise the vehicle while remaining responsible for resuming control upon request, fundamentally shifting their role from active control to supervisory monitoring [1, 2]. This transition introduces new human factors challenges related to maintaining attention and situational awareness during prolonged automation [3–5]. Drowsiness during supervisory automation has been shown to critically impair takeover performance and reduce safety margins, underscoring the need for reliable driver state monitoring in Level 3 systems [6–8]. Eye- and gaze-based approaches offer a non-intrusive means of continuously monitoring driver state and have been widely applied for drowsiness recognition in both manual and automated driving [9–11, 13]. However, prior studies offer limited quantitative guidance for selecting non-redundant gaze features and suitable temporal windows for real-time monitoring in supervisory driving [4, 14, 15]. The present work addresses these gaps by systematically identifying core gaze-based features and an optimal temporal window using a multi-stage feature selection procedure and cross-window reliability analysis.

2 Related Work

Previous studies on driver drowsiness recognition have extensively employed eye- and gaze-based features, including PERCLOS, blink/eyelid characteristics, pupil responses, and gaze movement metrics [9–11, 13]. Recent eye-tracking studies have shown that increasing sleepiness is associated with degraded situation awareness and altered visual attention allocation under automated driving conditions [16, 17]. Gaze-based features such as fixation duration and gaze dispersion have been found to sensitively reflect changes in drivers' cognitive engagement across different traffic and automation contexts [18]. Beyond detecting drowsiness or distraction, prior work has emphasized predicting driver readiness by linking cognitive state monitoring with in-vehicle decision support systems, underscoring the importance of anticipatory state estimation in automated driving [19]. To improve recognition performance,

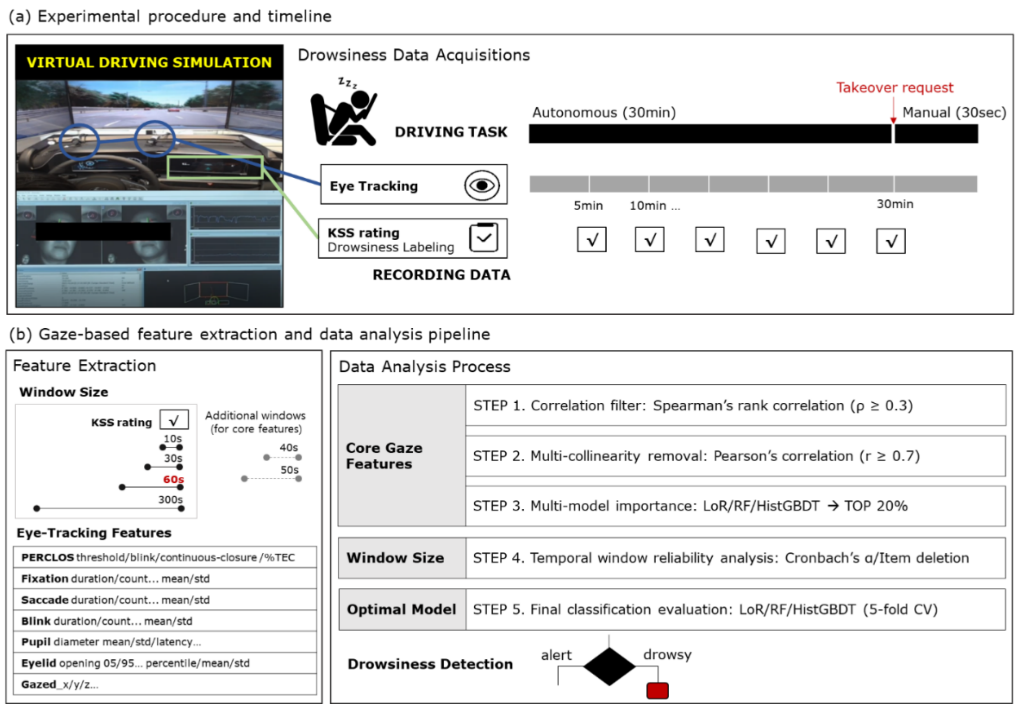

Figure 1: Overview of the proposed gaze-based drowsiness recognition framework

many approaches combine multiple gaze-based features and apply machine-learning or deep-learning methods for classification [14, 15]. However, such methods often rely on large feature sets without systematically addressing redundancy or multicollinearity among features, which can complicate system design and reduce interpretability [13]. Furthermore, the temporal window length critically affects feature stability and real-time applicability. However, most studies rely on fixed window lengths chosen empirically, without quantitatively evaluating representativeness or reliability across different windows. For example, Massoz et al. compared drowsiness-related facial features across short temporal windows ranging from 10 s to 60 s, demonstrating that window length influences the characterization of driver drowsiness [14].

3 Experimental Methods

3.1 Experimental Environment and Data Collection

To emulate a Level 3 automated driving environment, we built a driving simulator using production components from a commercial passenger vehicle (G model, H company). The driving scenario, implemented with the CARLA open-source platform, simulated a six-lane highway with an average speed of 100 km/h, allowing approximately 30 minutes of continuous automated driving. One hundred participants aged 20-50 years took part in the experiment, which was approved by the Institutional Review Board (No. N01-202407-01-013). Participants remained in fully automated mode and performed only a supervisory role without manual control inputs until a take-over request (TOR) was issued at the end of the automated driving period, after which they resumed manual driving. Eye-tracking data were continuously recorded at 60 Hz using a Smart Eye Pro system, and subjective drowsiness was assessed every five minutes using the Karolinska Sleepiness Scale (KSS; 1–9) [20]. Figure 1(a) summarizes the Level 3 experimental procedure and timeline, including the automated driving period, repeated KSS assessments, and the final take-over request, while Figure 1(b) illustrates the multi-stage analysis pipeline for multi-window gaze feature extraction, feature selection, window reliability assessment, and drowsiness classification.

3.2 Eye-tracker-based Feature Extraction and Window Configuration

For analysis, the five-minute interval immediately preceding each KSS rating (block5min) was defined as the primary analysis unit, within which 46 gaze-related features were computed. Each metric was calculated over four temporal windows (10, 30, 60, and 300 seconds) immediately before each KSS rating, enabling comparison of feature representativeness across temporal scales. In total, this yielded 46 features: 18 PERCLOS-based measures and 28 eye-tracking-based measures. PERCLOS features were derived from bilateral eye openness using multiple closure thresholds (70%, 75%, 80%), temporal windows (10, 30, 60, 300 seconds), and calculation schemes, including threshold-based PERCLOS, blink-based closure, continuous closure with minimum duration constraints, and percentage of time in eye closure (%TEC), yielding multiple PERCLOS variants from combinations of thresholds, windows, and closure definitions. Eye-tracking-based features captured blink and fixation characteristics, saccadic eye movements, gaze dispersion, and pupil responses, reflecting behavioral and physiological aspects such as frequency, duration, variability, and spatial spread to characterize reduced gaze stability and decreased arousal during early stages of drowsiness in Level 3 supervisory driving.

3.3 Core Feature Selection and Classification Analysis

To identify core gaze features, nonparametric correlations between KSS ratings and gaze features were first quantified using Spearman's rank correlation coefficient, and features with $\rho \geq 0.3$ ($p<0.001$) were retained as candidates associated with drowsiness. Given the design-oriented goal of screening for practically useful predictors rather than conducting strict hypothesis tests, these thresholds were treated as descriptive criteria for feature selection, and robustness was instead evaluated through cross-validated classification performance. Because multiple windows were obtained from each participant over time, the resulting p-values should therefore be interpreted as descriptive indicators rather than as the basis for strong inferential claims. To mitigate multicollinearity, we identified pairs of features with Pearson correlations $r \geq 0.7$ and, within each highly correlated pair, retained only the feature with the stronger association with KSS. The final non-collinear feature set was then used to classify drowsy (KSS>7) versus alert (KSS<7) states using three classifiers: Logistic Regression (LoR), Random Forest (RF), and Histogram Gradient Boosting Decision Tree (HistGBDT), and model-based feature importance was computed to select the final core metrics as those consistently showing high importance across models. To evaluate the representativeness of temporal windows for these core features, reliability across the four windows (300, 60, 30, and 10 seconds) was assessed using Cronbach's α and item-deletion analyses, and Pearson correlations and variance inflation factors (VIFs) were examined to confirm the absence of multicollinearity across windows; additional intermediate windows of 40 and 50 seconds were also computed for the four core gaze features to probe the transition between 30- and 60-second aggregation lengths. Finally, drowsiness classification performance was evaluated using 5-fold cross-validation with the LoR, RF, and HistGBDT models, with accuracy, sensitivity, specificity, and area under the ROC curve (AUC) as performance metrics. Overfitting was assessed using learning curves, training–validation performance differences(ΔAccuracy<5%), and temporal window reduction experiments. To further ensure valid evaluation, all cross-validation folds were constructed on a participant-wise basis so that all windows from a given participant appeared in only one fold, preventing leakage of person-specific patterns between training and test sets.

4 Results

4.1 Feature Selection and Reduction

To identify features associated with drowsiness, Spearman's rank correlation analysis was conducted between KSS scores and the extracted gaze-based features. After correlation filtering and multicollinearity removal, ten features were retained: three PERCLOS-based features (block5min_cont_perclos500_70, block5min_perclos70, and last60_cont_perclos500) and seven gaze-based features, including pupil variability, mean fixation duration, fixation time ratio, eyelid openness, and vertical gaze dispersion (Table 1). This reduced set preserved relevance to drowsiness while minimizing redundancy and was used as input for subsequent multi-model analysis.

4.2 Multi-model Analysis and Core Metrics Identification

Using the reduced set of ten features from the previous analysis, a multi-model classification analysis was conducted to distinguish between drowsy (KSS>7) and alert (KSS<7) states. To avoid model-specific bias in metric selection, three classifiers - LoR, RF, and HistGBDT - were applied under identical conditions. All models were trained using the same input features, and model-based feature importance was computed for each classifier. The comparison of feature importance across models revealed that while some features showed high importance only in specific classifiers, a subset of gaze-based features consistently exhibited substantial importance across multiple models (Table 2). Based on this consistency criterion, four core features were identified: pupil variability (pupil_std), mean fixation duration (fix_dur_mean), mean eyelid openness (lid_mean), and vertical gaze dispersion (gazed_std_y). These core features capture complementary aspects of drowsiness-related behavior. Pupil variability reflects changes in autonomic arousal, mean fixation duration indicates shifts in gaze maintenance, eyelid openness indexes early signals of eye closure, and vertical gaze dispersion reflects reduced gaze stability. These four features thus provide a compact and complementary summary of the physiological and behavioral processes underlying drowsiness-related gaze behavior.

4.3 Window Reliability and Representativeness Analysis

Reliability analysis was conducted to evaluate the representativeness of different temporal windows for the four core gaze features (pupil_std, fix_dur_mean, lid_mean, and gazed_std_y). Overall internal consistency across windows was high (Cronbach's $\alpha = .76 \sim .97$), indicating that the core features were generally stable across temporal scales. Table 3 summarizes Cronbach's α and ICC values for the four core features across all six windows. In particular, lid_mean showed very high consistency (α =.972) and remained stable even at the 10-second window, whereas fix_dur_mean exhibited reduced consistency when the 10-second window was included, suggesting higher susceptibility to short-term fluctuations. Across features, the 60-second window showed the strongest alignment with the 300-second window (e.g., r=.812 for pupil_std; r=.814 for fix_dur_mean; r=.933 for lid_mean; r=.857 for gazed_std_y), and removing the 60-second window produced the largest reduction in overall reliability based on item-deletion analysis (Figure 2). In addition to the 10-, 30-, 60-, and 300-second windows, intermediate 40- and 50-second windows were also examined for the four core features. For pupil variability, eyelid openness, and vertical gaze dispersion, the 40- and 50-second windows showed very high correlations with the 60-second window ($r \geq .969$ for pupil_std; $r \geq .984$ for lid_mean; $r \geq .960$ for gazed_std_y) and similarly high Cronbach's α values, indicating a plateau in representativeness between 40 and 60 seconds, with 30-second windows remaining slightly less stable

Table 1: Feature selection and reduction based on correlation and multicollinearity analysis

Feature	Category	Window	ρ	Feature	Category	Window	ρ
block5min_cont_perclos500_70_hz60*	PERCLOS	300s	0.387	block5min_pupil_std*	Pupil	300s	0.517
block5min_cont_perclos400_70_hz60	PERCLOS	300s	0.381	block5min_fix_time_ratio*	Fixation	300s	−0.508
block5min_cont_perclos500_70_hz24	PERCLOS	300s	0.380	block5min_gazed_std_y*	Gaze	300s	0.502
block5min_cont_perclos400_70_hz24	PERCLOS	300s	0.376	block5min_fix_time_sec*	Fixation	300s	−0.499
block5min_cont_perclos500_70_hz10	PERCLOS	300s	0.372	block5min_pupil_p05*	Pupil	300s	−0.486
block5min_perclos70_hz6*	PERCLOS	300s	0.370	last60_pupil_std	Pupil	60s	0.428
block5min_tec_pctwin70_hz6	PERCLOS	300s	0.370	block5min_lid_mean*	Eyelid	300s	−0.419
block5min_cont_perclos500_70_hz6	PERCLOS	300s	0.368	last60_gazed_std_y*	Gaze	60s	0.417
block5min_cont_perclos400_70_hz10	PERCLOS	300s	0.365	block5min_lid_p05	Eyelid	300s	−0.400
last60_cont_perclos500_70_hz60*	PERCLOS	60s	0.342	block5min_lid_p50	Eyelid	300s	−0.392

Note. Features marked with an asterisk indicate the final features retained after collinearity analysis and used as inputs for subsequent classification and window reliability analyses

Table 2: Multi-model feature importance and identification of core gaze features

Feature	LoR (β)	RF	HistGBDT	Feature	LoR (β)	RF	HistGBDT
block5min_pupil_std*	0.734	0.036	0.038	block5min_lid_mean*	−0.681	0.022	0.038
block5min_fix_time_ratio*	1.012	0.032	0.035	block5min_gazed_std_y*	0.611	0.025	0.035
block5min_fix_dur_mean*	0.904	0.030	0.033	block5min_pupil_p95*	−0.598	0.021	0.033
block5min_pupil_p05	−0.702	–	0.031	block5min_gazed_std_z	–	0.020	0.031
block5min_pupil_f_std	–	0.026	0.029	block5min_sacc_rate_per_min*	−0.645	0.019	0.029

Note. Features marked with an asterisk indicate the four core gaze features selected for the subsequent window reliability analysis.

Table 3: Internal consistency (Cronbach's α) and intraclass correlation coefficients (ICC, single measure, absolute agreement) for the four core gaze features across six temporal windows (300, 60, 50, 40, 30, and 10 s)

Measure	Windows included	Cronbach's α	Cronbach's α range	ICC
Pupil variability (pupil_std)	300s	.961	.944 – .973	.795
	60s		(max. α =.973 excluding 10s)	
Mean fixation duration (fix_dur_mean)	50s	.876	.828 – .973	.539
	40s		(max. α =.973 excluding 10s)	
Eyelid openness (lid_mean)	30s	.987	.982 – .991	.927
	10s		(max. α =.991 excluding 10s)	
Vertical gaze dispersion (gazed_std_y)		.971	.959 – .979	.844
			(max. α =.979 excluding 10s)	

Note. For all measures, the highest Cronbach's α is obtained when the 10-second window is excluded, indicating that the shortest window contributes most to residual instability.

and 10-second windows clearly more affected by transient fluctuations. Together, these results indicate that the 60-second window offers a practical balance between temporal representativeness and real-time applicability for Level 3 supervisory driving.

4.4 Final Classification Performance

Final classification was conducted using the four core gaze features - pupil variability (pupil_std), mean fixation duration (fix_dur_mean), mean eyelid openness (lid_mean), and vertical gaze dispersion (gazed_std_y) - computed over the 60-second window (Table 4). Across all models, accuracy exceeded 0.75, indicating that even a compact feature set can provide reasonably reliable discrimination between alert (KSS<7) and drowsy (KSS>7) states in Level 3 supervisory driving. The Random Forest model achieved the best overall performance (Accuracy=0.781, AUC=0.832), with balanced sensitivity and specificity, while Logistic Regression and HistGBDT yielded slightly lower but similar accuracies, suggesting that nonlinear modeling yields only modest gains over simpler models in this configuration. The comparable performance across classifiers implies that the measure-window configuration itself - rather than

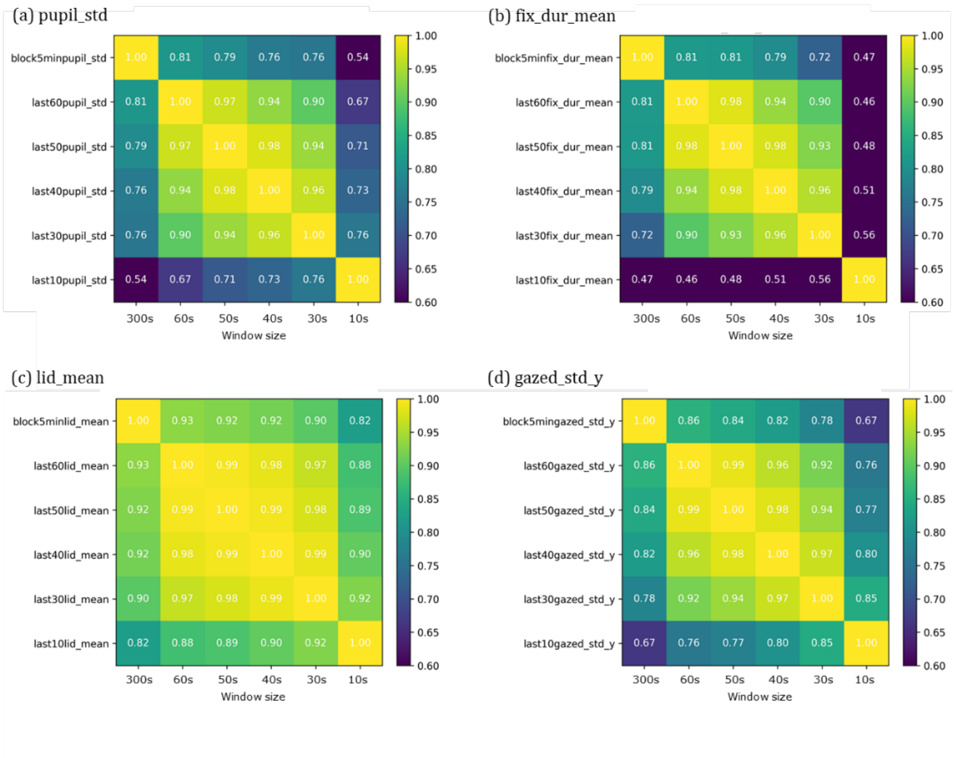

Figure 2: Cross-window correlations of core gaze features. Heatmaps show Pearson correlations between values computed over 10-, 30-, 40-, 50-, 60-, and 300-second windows for the four core gaze features: (a) pupil variability, (b) mean fixation duration, (c) mean eyelid openness, and (d) vertical gaze dispersion.

Table 4: Final drowsiness classification performance using the selected features and temporal window

Model	Accuracy	Balanced Acc	Sensitivity	Specificity	ROC AUC
LoR	0.751	0.734	0.721	0.747	0.801
RF	0.781	0.763	0.752	0.774	0.832
HistGBDT	0.776	0.758	0.747	0.769	0.826

a particular learning algorithm - is the primary driver of detection performance, supporting its suitability as a design-oriented baseline for gaze-based drowsiness monitoring.

5 Discussions

This study adopts a design-oriented perspective on gaze-based drowsiness recognition for SAE Level 3 automated driving, examining which features to prioritize and how to configure temporal windows. Prior studies have shown that eye- and gaze-based cues are effective for drowsiness detection and have often relied on short temporal windows for real-time monitoring [14, 15]. Much of this work, however, has emphasized boosting classification performance rather than explicitly testing whether short-term features reliably stand in for longer-term drowsiness states in supervisory driving contexts, making it difficult to derive concrete design guidance for Level 3 driver monitoring systems. When comparing multiple classification models, four gaze-based features - pupil variability, mean fixation duration, eyelid openness, and vertical gaze dispersion - consistently emerged as core indicators. Previous eye-tracking research has reported that pupil responses closely track changes in autonomic activity and cognitive workload during automated driving, while fixation duration and eyelid behavior reflect reduced cognitive engagement and physical signs of fatigue [16–19]. Neurophysiological work further suggests that pupillary variability reflects the dynamic balance between sympathetic and parasympathetic activity, with increased parasympathetic (vagal) modulation emerging during states of reduced arousal [20]. Taken together, these findings support the view that pupil-based features can capture early autonomic changes that precede shifts in gaze behavior and overt drowsiness-related movements, and the present results reinforce the value of a multidimensional representation of gaze behavior rather than reliance on a single cue such as eye closure

[13]. With respect to time scales, the 60-second window provided a practical balance between representativeness and responsiveness. Although short windows have been widely adopted in prior work for real-time detection [14, 15], our reliability analysis indicates that very short windows (10 s and 30 s) are more vulnerable to transient fluctuations. In contrast, the 60-second window aligns closely with longer windows while still being feasible for real-time deployment. Additional analyses with intermediate windows of 40 and 50 seconds support this interpretation, as windows between 40 and 60 seconds produced nearly indistinguishable reliability and cross-window correlations, suggesting a plateau in representativeness around this range. Within this plateau, the 60-second window provides a convenient design choice that balances stable estimation with implementation simplicity, while 30-second windows remain a viable option when shorter detection latency is prioritized. Notably, the similar classification performance observed across different models suggests that the proposed measure–window configuration is not tied to any specific learning algorithm but instead offers a model-agnostic baseline that can be integrated into a variety of detection pipelines. From a design perspective, this implies that system developers can focus on implementing robust sensing and feature computation while retaining flexibility in their choice of downstream classifiers. The study was conducted in a simulator-based environment and relied on subjective drowsiness ratings, which may limit generalizability [16]. In particular, KSS provides a coarse self-reported snapshot of sleepiness rather than a continuous behavioral performance measure, and simulator conditions may underestimate environmental perturbations present in real traffic. Recent work has emphasized the importance of multimodal sensing, cross-domain generalization, and adaptive personalization for improving the robustness of driver fatigue monitoring systems [22–24]. Building on this direction, future research will extend the proposed framework to real-world driving conditions and investigate how the recommended feature-window configuration behaves across diverse traffic, environmental, and driver populations, while incorporating additional physiological signals to evaluate and refine the identified design choices. Finally, while this work recommends a 60-second window as a strong default, shorter windows such as 30 seconds may remain preferable in applications where detection latency is paramount, making the exploration of such application-specific trade-offs an important direction for follow-up studies.

6 CONCLUSIONS

This study identified key design principles for gaze-based drowsiness recognition in SAE Level 3 automated driving. By systematically analyzing a large set of gaze- and PERCLOS-based features, this study demonstrates that reliable drowsiness recognition can be achieved with four core gaze features and a 60-second temporal window, offering concrete guidance on which measures to use and over what time span when designing real-time supervisory monitoring systems. Future work will investigate the temporal dynamics preceding takeover requests by examining how changes in autonomic activity (pupil variability), cognitive engagement (fixation duration), and physical eye behavior (eyelid movements) unfold over time. In particular, the aim is to characterize the progression from early pupil-based autonomic changes, through shifts in fixation patterns, to overt eyelid movements that mark pronounced drowsiness. Such temporal modeling, combined with extension of the framework to real-world driving data and multimodal physiological signals, may enable earlier and more reliable anticipation of drowsiness-related takeover failures in naturalistic supervisory driving conditions.

Acknowledgments

This work is supported by Korea Evaluation Institute of Industrial Technology (KEIT) grant funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea) (No. 20024892, Development of a heterogeneous sensor fusion module and integrated system for determining driver's physical and cognitive state; No. 20018248, Development of safety of the intended functionality from insufficiency of perception and decision making).

References

[1] SAE International. 2021. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE Standard J3016.
[2] Michael A. Gerber, Ronald Schroeter, and Bonnie Ho. 2023. A human factors perspective on how to keep SAE Level 3 conditional automated driving safe. Transportation Research Interdisciplinary Perspectives 22 (2023), 100959. https://doi.org/10.1016/j.trip.2023.100959
[3] Mica R. Endsley. 2017. From here to autonomy: Lessons learned from human–automation research. Human Factors 59, 1 (2017), 5–27. https://doi.org/10.1177/0018720816681350
[4] Christian Gold, Daniel Damböck, Laura Lorenz, and Klaus Bengler. 2013. "Take over!" How long does it take to get the driver back into the loop? Human Factors 55, 2 (2013), 193–205. https://doi.org/10.1177/0018720812463816
[5] Hyunsuk Kim, Woojin Kim, Jungsook Kim, Seung-Jun Lee, Daesub Yoon, Oh-Cheon Kwon, and Cheong Hee Park. 2023. Study on the Take-over Performance of Level 3 Autonomous Vehicles Based on Subjective Driving Tendency Questionnaires and Machine Learning Methods. ETRI Journal 45, 1 (2023), 75–92. https://doi.org/10.4218/etrij.2021-0241
[6] Erin E. Flynn-Evans, Lisa R. Wong, Yuki Kuriyagawa, *et al.* 2021. Supervision of a self-driving vehicle unmasks latent sleepiness relative to manually controlled driving. Scientific Reports 11 (2021), 18530. https://doi.org/10.1038/s41598-021-92914-5
[7] Yusuke Wu, Koji Kihara, Kazuki Hasegawa, Yuki Takeda, Taro Sato, Masao Akamatsu, and Shoji Kitazaki. 2020. Age-related differences in effects of non-driving related tasks on takeover performance in automated driving. Journal of Safety Research 72 (2020), 231–238. https://doi.org/10.1016/j.jsr.2019.12.019
[8] Haotian Pan, David B. Logan, Amanda N. Stephens, William Payre, Ying Wang, Zhi Peng, Yilin Qin, and Stuart Koppel. 2025. Exploring the effect of driver drowsiness on takeover performance during automated driving: An updated literature review. Accident Analysis & Prevention 216 (2025), 108023. https://doi.org/10.1016/j.aap.2025.108023
[9] David F. Dinges and Roger Grace. 1998. PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance. FHWA Report No. FHWA-MCRT-98-006
[10] Reinhard Schleicher, Nadine Galley, Sabine Briest, and Lutz Galley. 2008. Blink and eyelid parameters as indicators of driver fatigue. Applied Ergonomics 39, 6 (2008), 798–805. https://doi.org/10.1016/j.apergo.2008.01.006
[11] Qiang Ji, Zhiwei Zhu, and Peilin Lan. 2004. Real-time nonintrusive monitoring and prediction of driver fatigue. IEEE Transactions on Vehicular Technology 53, 4 (2004), 1052–1068. https://doi.org/10.1109/TVT.2004.832447
[12] Nadja Schömig, Verena Hargutt, Alexander Neukum, Ina Petermann-Stock, and Ines Othersen. 2015. The interaction between highly automated driving and the development of drowsiness. Procedia Manufacturing 3 (2015), 6652–6659. https://doi.org/10.1016/j.promfg.2015.07.879
[13] Tetsuo Abe. 2023. PERCLOS-based technologies for detecting drowsiness: Current evidence and future directions. Sleep Advances 4, 1 (2023), zpad006. https://doi.org/10.1093/sleepadvances/zpad006
[14] Quentin Massoz, Justine G. Verly, and Marc Van Droogenbroeck. 2018. Multi-timescale drowsiness characterization based on a video of a driver's face. Sensors 18, 9 (2018), 2801. https://doi.org/10.3390/s18092801
[15] Guofa Li and Wan-Young Chung. 2022. Detection of driver drowsiness using eye movements and deep learning. IEEE Sensors Journal 22, 1 (2022), 404–415. https://doi.org/10.1109/JSEN.2021.3125895

[16] Stefan M. Castritius, Martin Braun, and Klaus Bengler. 2021. Driver situation awareness and perceived sleepiness during truck platoon driving: Insights from eye-tracking data. Transportation Research Part F 78 (2021), 1–15. https://doi.org/10.1016/j.trf.2021.02.012

[17] Vitor Silva Gomes, Ana Margarida Madureira, and Filipe Duarte. 2024. The effect of sleepiness on situation awareness: A scoping review. Safety Science 169 (2024), 106309. https://doi.org/10.1016/j.ssci.2023.106309

[18] Aurélien Halin, Aurélie Pêcher, and Thierry Baccino. 2024. Gaze-based indicators of driver cognitive distraction: Effects of different traffic conditions and adaptive cruise control use. Transportation Research Part F 97 (2024), 102343. https://doi.org/10.1016/j.trf.2023.102343

[19] Mi Chang, Eun Hye Jang, Woojin Kim, Daesub Yoon, and Do Wook Kang. 2025. Driver readiness prediction: Bridging cognitive distraction monitoring and in-vehicle decision support systems. Decision Support Systems 199 (2025), 1–13. https://doi.org/10.1016/j.d

[20] Torbjörn Åkerstedt and Mats Gillberg. 1990. Subjective and objective sleepiness in the active individual. International Journal of Neuroscience 52, 1–2 (1990), 29–37. https://doi.org/10.3109/00207459008994241

[21] Andreas Schumann, Sebastian Kietzer, Jonas Ebel, and Karl-Jürgen Bär. 2020. Sympathetic and parasympathetic modulation of pupillary unrest. Frontiers in Neuroscience 14 (2020), 178. https://doi.org/10.3389/fnins.2020.00178

[22] Shuang Cao, Peng Feng, Wei Kang, *et al.* 2025. Optimized driver fatigue detection method using multimodal neural networks. Scientific Reports 15 (2025), 12240. https://doi.org/10.1038/s41598-025-86709-1

[23] Vishnu Radhakrishnan *et al.* 2023. Using pupillometry and gaze-based metrics for understanding drivers' mental workload during automated driving. Transportation Research Part F 93 (2023), 101868. https://doi.org/10.1016/j.trf.2023.02.015

[24] Aravinth S. S., G. M. Nagamani, C. K. Kumar, *et al.* 2025. Dynamic cross-domain transfer learning for driver fatigue monitoring: Multi-modal sensor fusion with adaptive real-time personalization. Scientific Reports 15 (2025), 15840. https://doi.org/10.1038/s41598-025-92701-6