

Predicting Reaction Time Under Level 2 Assisted Driving and Cognitive Load: A Sensitivity Analysis of Eye-Tracking Features Using XGBoost and Explainable AI

Hao Qin¹, Rafael C. Gonçalves¹, Courtney M. Goodridge^{1,2}, Natasha Merat¹

¹ Institute for Transport Studies, University of Leeds, UK

² School of Psychology, University of Leeds, UK



Co-funded by the
European Union under
Horizon 2020 programme
Grant Agreement No 101006664

01 Background

02 Methods

03 Results

04 Conclusions

1.1 The Growing Demand for Driver Monitoring Systems

➤ **Regulatory requirements**

Driver-monitoring systems (DMS) are increasingly expected in new vehicles (*Euro NCAP, 2023; EU Commission, 2022*).

➤ **Prevalence of level 2 assisted driving**

While using L2 AD (SAE, 2021), drivers are more inclined to take up non-driving-related tasks (NDRTs; *Klauer et al., 2023; Mueller et al., 2024*) yet the system still requires them to monitor the road and system status.



1.2 Why Cognitive Load Must Be Considered

➤ **Cognitive load is:**

The amount of a driver's finite cognitive-control resources occupied by a task (*Engström et al., 2013; ISO, 2016*).

➤ **Cognitive control is:**

A set of mental functions that regulate attention and behaviour when automatized responses are prone to fail. $\approx$ executive function (inhibition, working memory, planning, decision-making). (Diamond, 2013; Engström et al., 2013)

➤ **Cognitive control hypothesis**

Cognitive load from non-visual impairs performance that relies on cognitive control, such as brake response based on interpreting cues (e.g., downstream traffic, braking light), while leaving automatized responses unaffected or even improved (*Engström et al., 2017*).



1.2 Why Cognitive Load Must Be Considered

➤ ***The level-2 caveat***

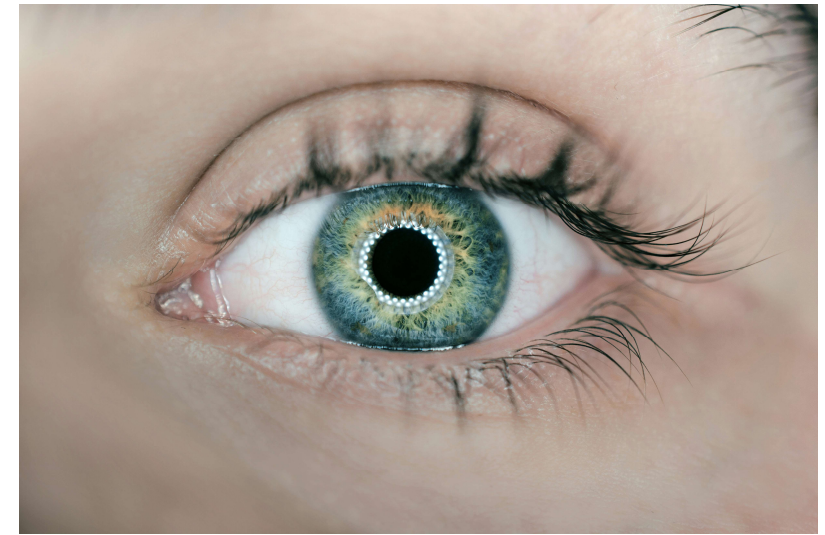
In L2 AD, evidence mostly shows limited effects of cognitive load on takeover/hazard response (*Choi et al., 2020; Goodridge et al., 2026; Louw et al., 2017*), but this may reflect scenario bias (e.g., alerted takeover with takeover request; lead-vehicle braking events; *Qin et al., 2026*).



Overall, monitoring cognitive load and anticipating its **detrimental effects** on **load-sensitive driver performance** is important in both manual and Level 2 assisted driving.

1.3 Related Work – Eye-Tracking Monitoring of Cognitive Load

- **Eye-tracking features** are established as sensitive to cognitive load (e.g., fixation dispersion/entropy, pupil dilation), but a DMS must also know **how these map onto safety-critical driver performance** (e.g., reaction time to hazards) to inform intervention;
- However, the value of eye-tracking features for representing load states to *predict performance* (e.g., *reaction time*) remains barely examined;
- Studies that do examine them have not tested **load-sensitive response** (e.g., driver-initiated response without takeover requests during L2 AD, hazard response relying on anticipation rather than looming).



1.4 Related Work – Divergence in Modelling Approach

- Prior work spans **linear regression** (interpretable but limited to linear effects) and **machine learning** that captures non-linear feature interactions (e.g., XGBoost; Chen et al., 2024; Gonçalves et al., 2025; Liang et al., 2007; Wu et al., 2021);
- Reaction-time prediction is framed either as **classification**, flagging "late" responses (e.g., $\geq$ 85th-percentile RT), or as **regression** predicting continuous RT.



1.4 Research Questions

- **RQ1:** Which eye-tracking indicators of cognitive load are predictive of drivers' reaction time, and does including automation context improve the prediction?
- **RQ2:** Do non-linear modelling approaches predict reaction time more accurately than linear ones?
- **RQ3:** Do the predictive indicators, and the relative performance of modelling approaches, differ between classification and regression formulations of reaction-time prediction?



01 Background

02

Methods

03 Results

04 Conclusions

2.1 Dataset: Urban Hazard Responses Under Cognitive Load and L2 AD

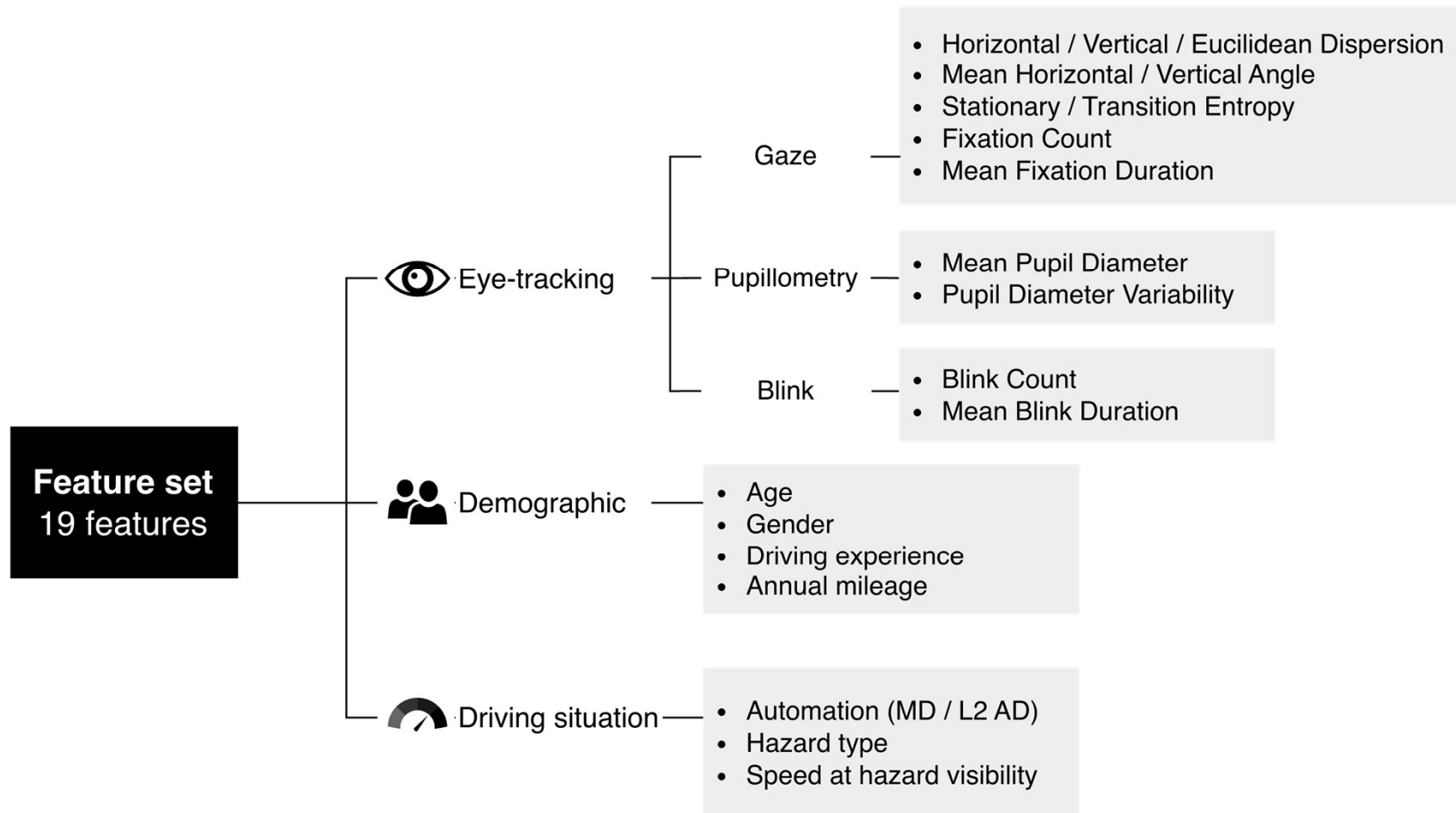
- **Participants:** 46 UK drivers (25 male, 21 female), aged 23–64 (M = 39.4 years)
- **2 x 2 Within-Subject Design:** Automation (Manual Driving vs. L2 AD) × Cognitive Load (No Task vs. 2-Back)
- **Hands-Off L2 AD:** Cruise control with lane-keeping assistance; the system could not detect or respond to hazards
- **Scenario:** Daytime urban driving (30 mph limit) with three types of critical hazards, each paired with a non-critical event to ensure uncertainty



Silent failures and driver-initiated responses to predictable hazards (Qin et al., 2026)



2.2 Machine Learning - Feature Set



2.3 Machine Learning – Eye-Tracking Time Windows

Data Preparation



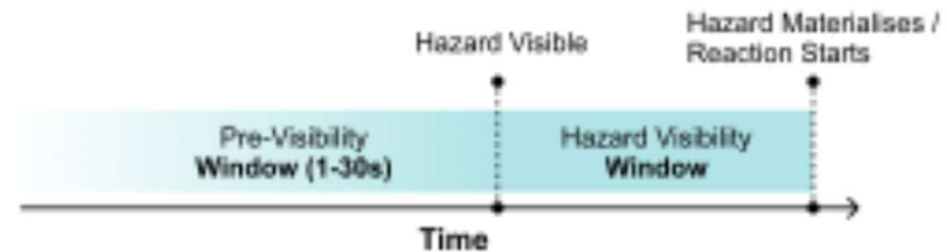
Experimental dataset

- Eye-tracking, demographic, and situational data

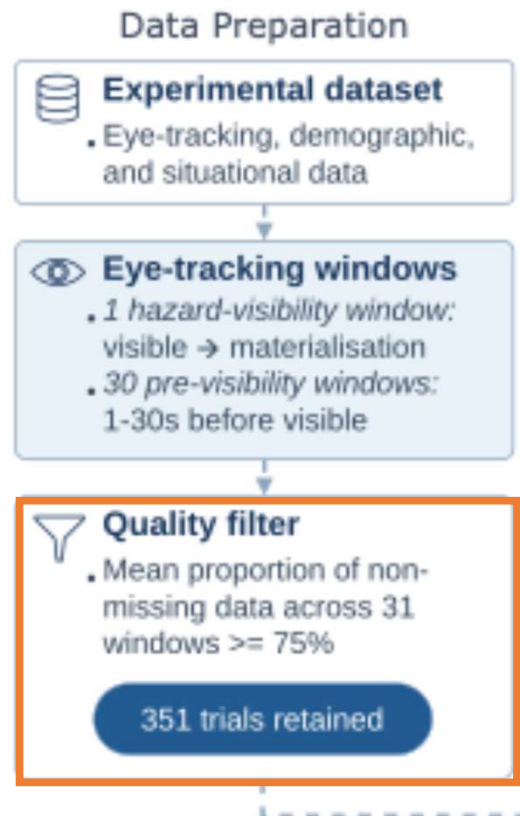


Eye-tracking windows

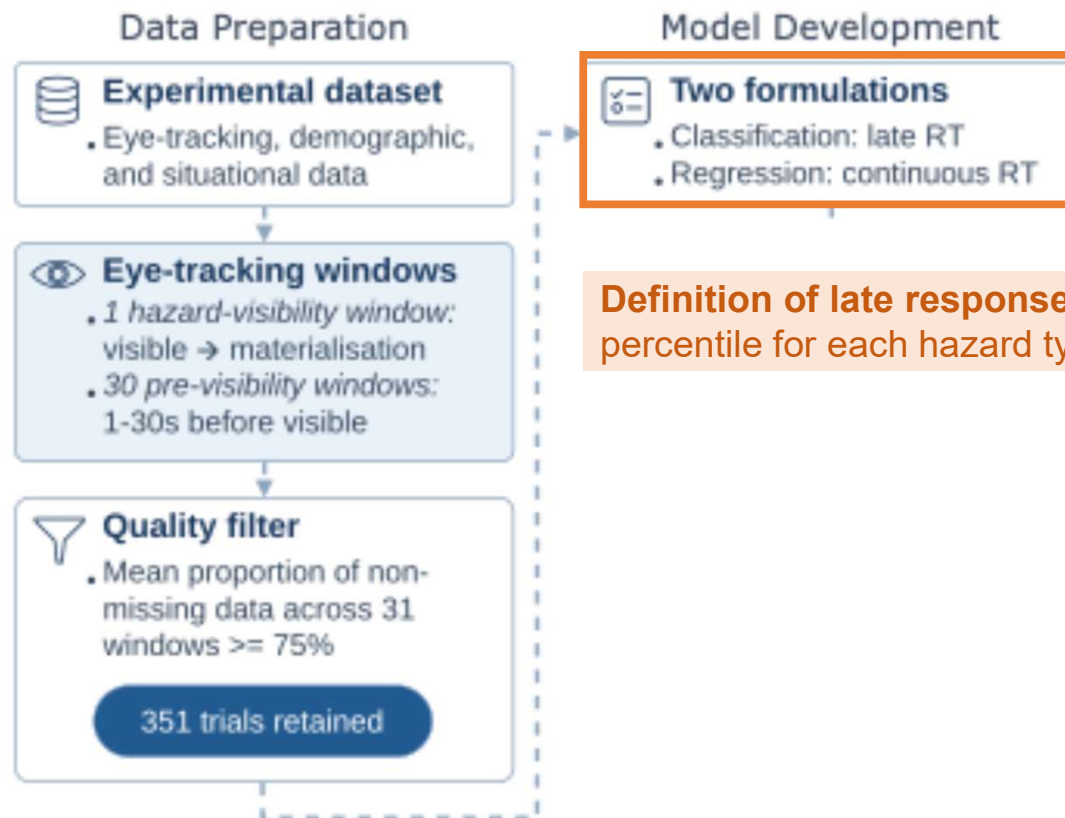
- 1 hazard-visibility window: visible → materialisation
- 30 pre-visibility windows: 1-30s before visible



2.3 Machine Learning – Eye-Tracking Quality Filter

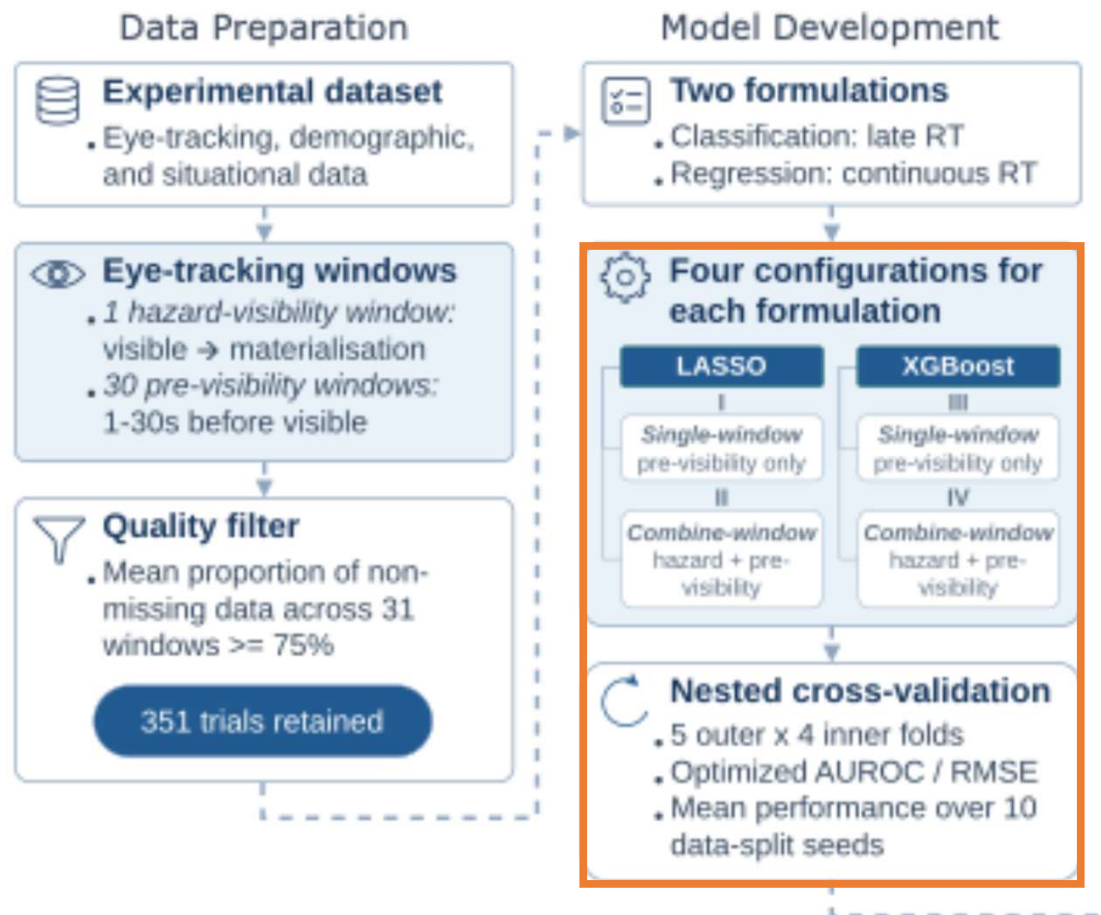


2.3 Machine Learning – Task Formulation

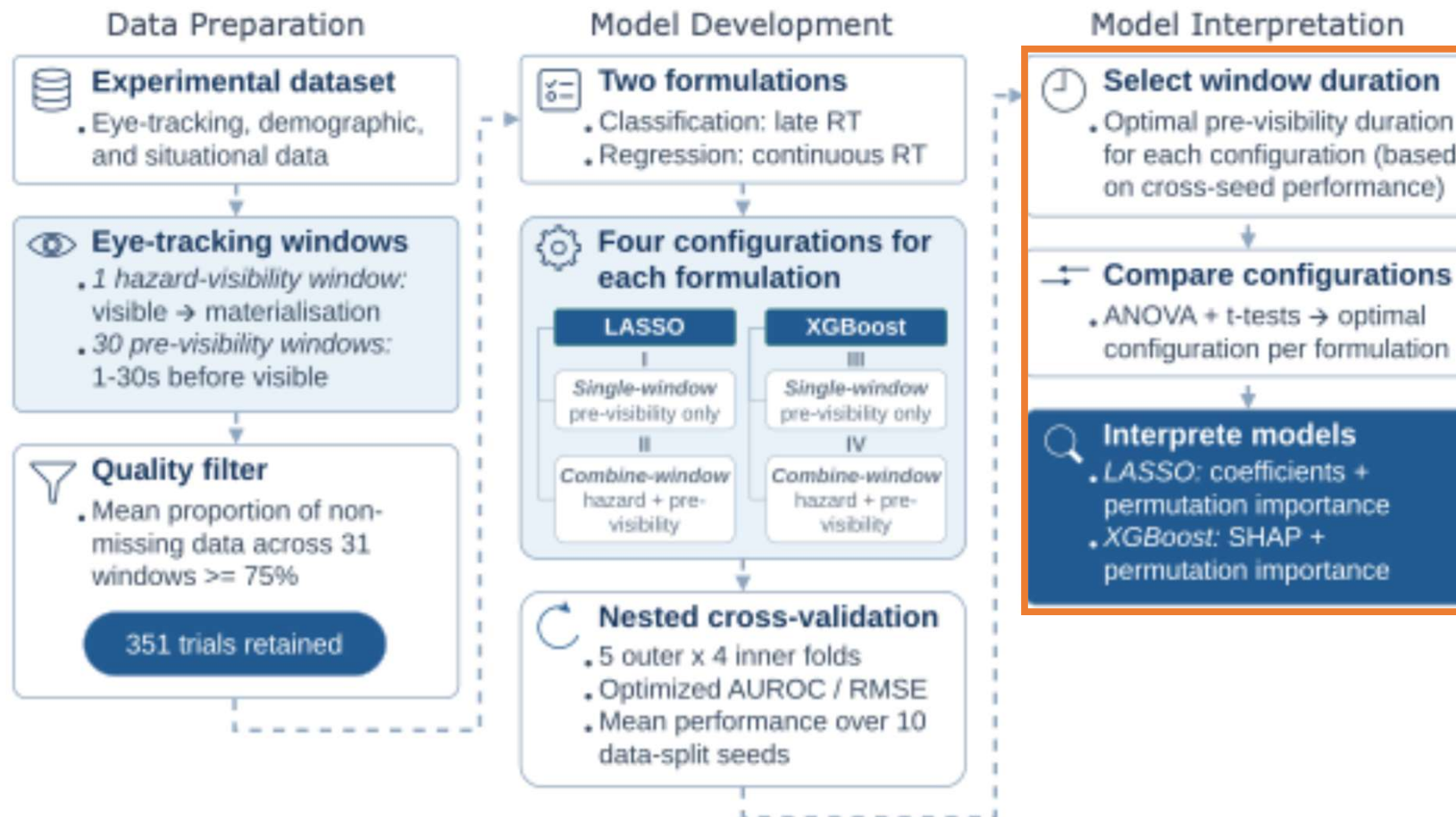


Definition of late response: RT at or above the 85th percentile for each hazard type.

2.3 Machine Learning – Modelling



2.3 Machine Learning – Evaluation and Interpretation



01 Background

02 Methods

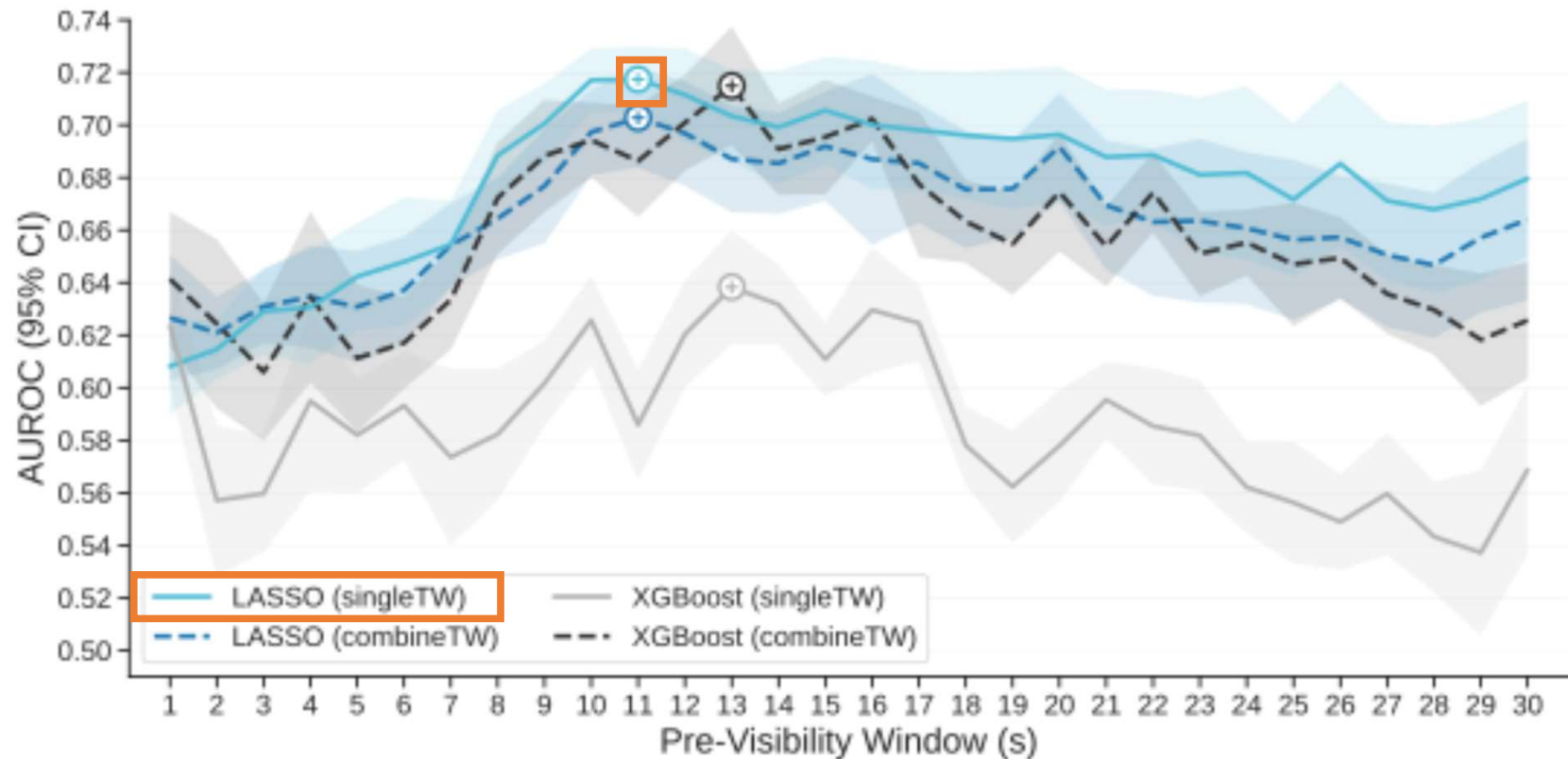
03

Results

04 Conclusions

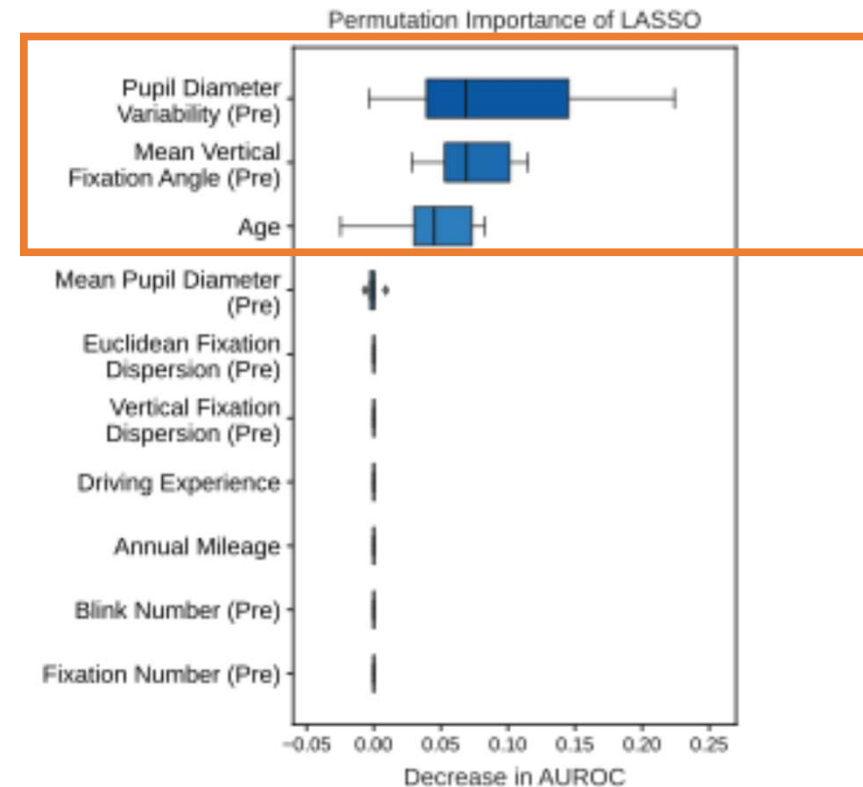
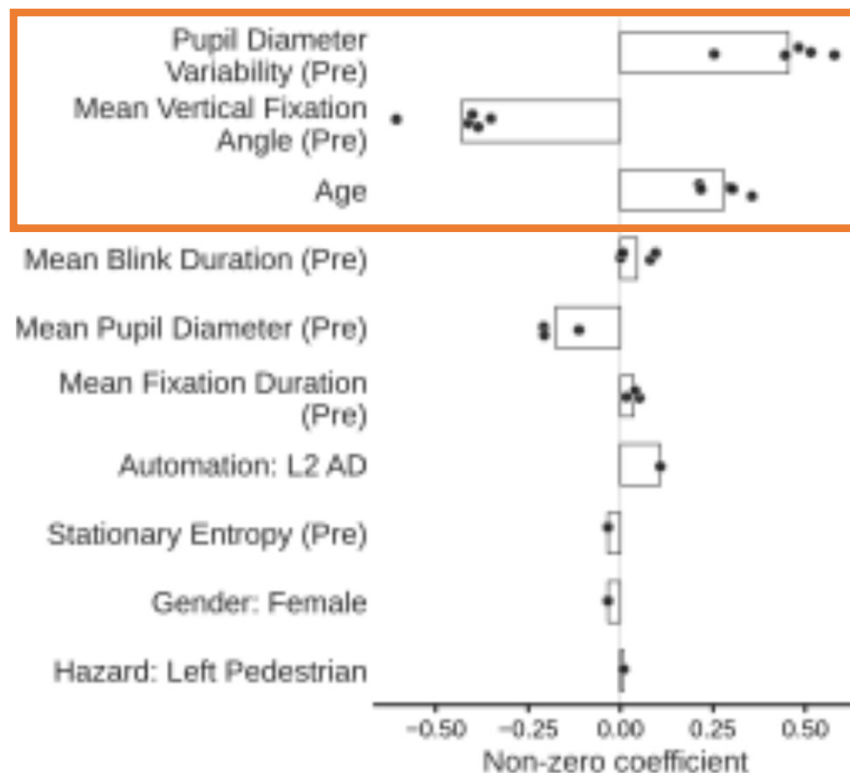
3.1 Classification Performance - AUROC

- ***Simplest model wins:*** LASSO with 11s pre-visibility window (AUROC=0.717);
- Pre-visibility window flagged “late” response in advance even before hazards became visible;



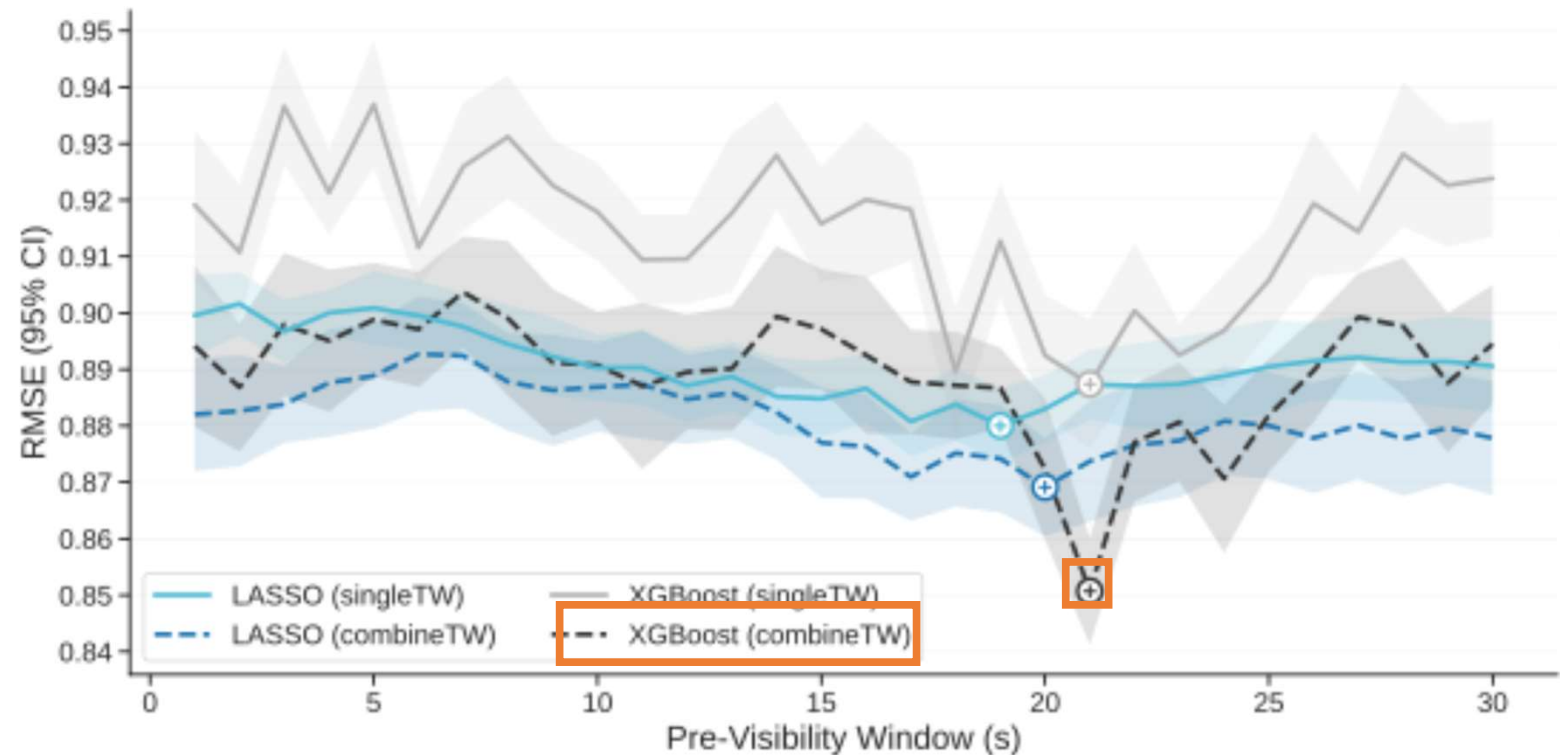
3.2 Model Interpretation – LASSO Coefficients and PFI

➤ **Influential features:** pupil diameter variability, mean vertical fixation angle, and age.



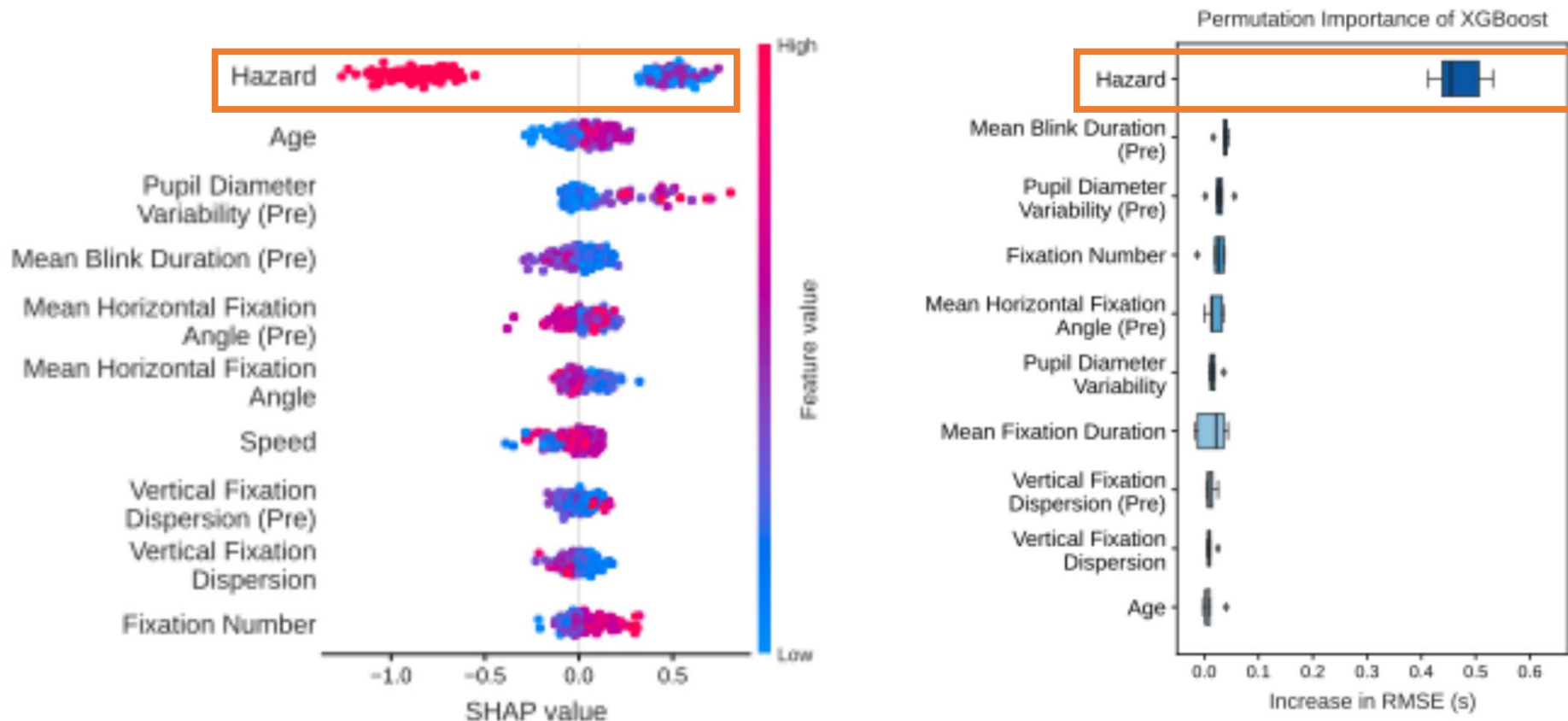
3.3 Regression Performance - RMSE

- **Complex model wins:**
XGBoost with a 21 s
pre-visibility window
combined with the
hazard-visibility
window
(RMSE=0.851s)



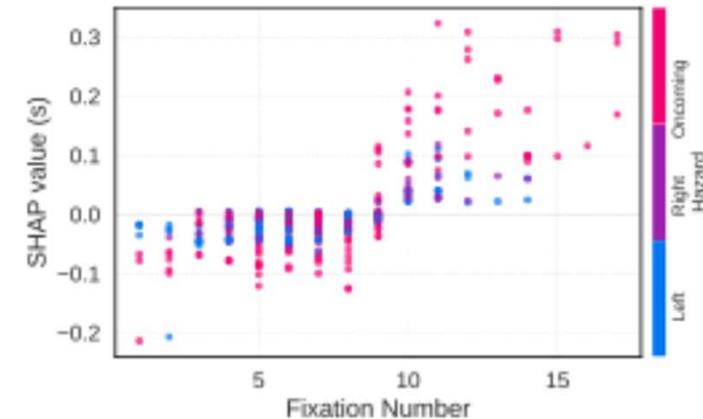
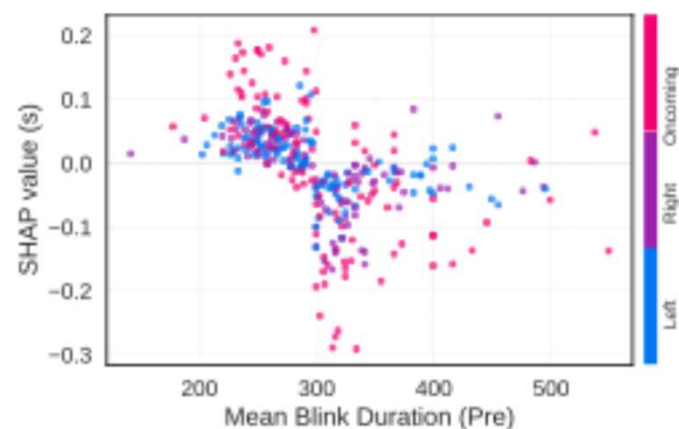
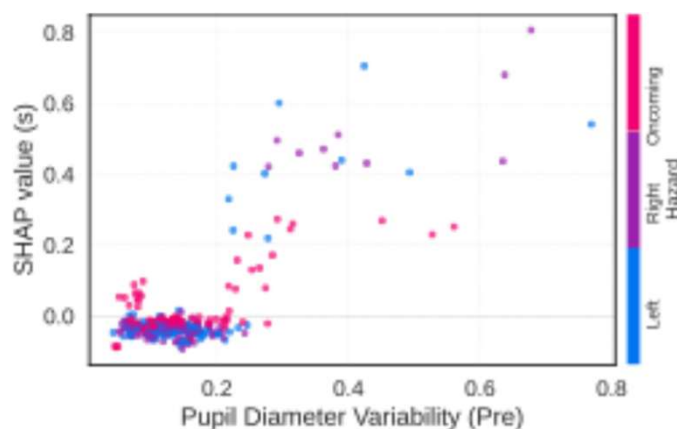
3.4 Model Interpretation – XGBoost SHAP and PFI

➤ *Hazard type emerged as the single dominant feature, far ahead of all other features.*



3.4 Model Interpretation – SHAP Feature Interaction

- Hazard type was also the **dominant interaction feature** with all others. (e.g. stronger effects of pupil diameter variability for pedestrian hazards, but stronger effects of blink duration and fixation number for vehicle hazards)



01 Background

02 Methods

03 Results

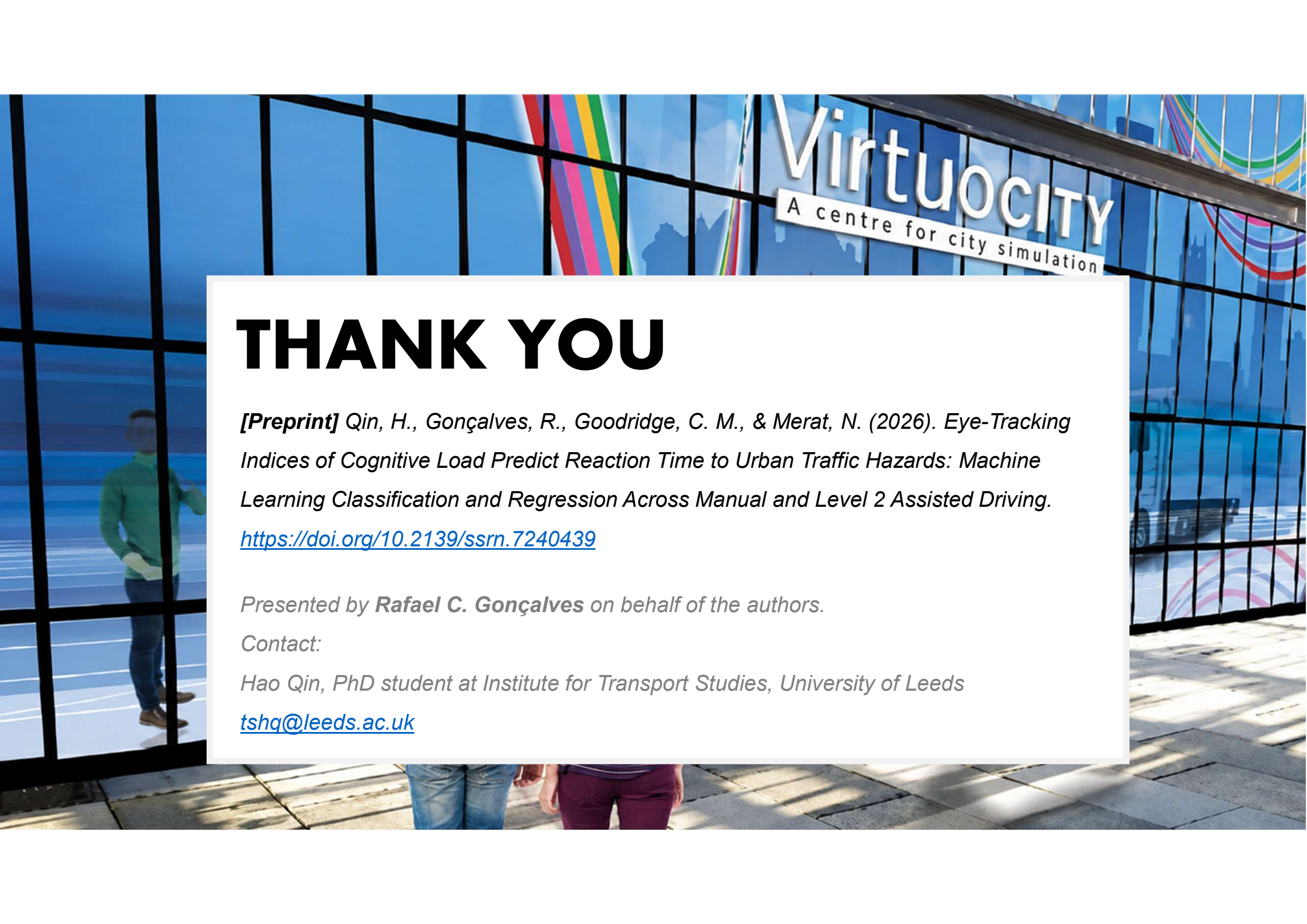
04 Conclusions

4.1 Conclusions

- Eye-tracking load indices **predicted "late" responses even before the hazards became visible** (LASSO with pre-visibility window reached AUROC = 0.717);
- Pupil diameter variability, mean vertical fixation angle, and age were the key predictors of late responses;
- Reaction-time **regression needed a more complex model than classification** (XGBoost beat LASSO), and was dominated by hazard type (event criticality) over other features;
- A DMS could utilize eye-tracking features **to anticipate the performance-decrement** risk associated with cognitive load states;
- Automation context (manual vs. L2 AD) added little predictive value — a single model may generalise across both modes;
- However, the current results worth verification with larger samples, real-driving data and explicit modelling of interactions.

References

- Choi, D., Sato, T., Ando, T., Abe, T., Akamatsu, M., & Kitazaki, S. (2020). Effects of cognitive and visual loads on driving performance after take-over request (TOR) in automated driving. *Applied Ergonomics*, 85, 103074.
- Engström, J., Markkula, G., Victor, T., & Merat, N. (2017). Effects of Cognitive Load on Driving Performance: The Cognitive Control Hypothesis. *Human Factors*, 59(5), 734–764.
- Engström, J., Monk, C. A., Hanowski, R. J., Horrey, W. J., Lee, J. D., McGehee, D. V., Regan, M., Stevens, A., Traube, E., Tuukkanen, M., Victor, T., & Yang, C. Y. D. (2013). *A conceptual framework and taxonomy for understanding and categorizing driver inattention. European Commission. Directorate General for Communications Networks, Content & Technology.*
- Euro NCAP. (2023). *Assessment Protocol – Safety Assist, Safe Driving: Implementation 2023. Version 10.3. Retrieved from:*
- European Commission. (2022). *Road safety thematic report – Driver distraction. European Road Safety Observatory. Brussels, European Commission, Directorate General for Transport.*
- Goodridge, C. M., Gonçalves, R. C., Arabian, A., Horrobin, A., Solernou, A., Lee, Y. T., Bruneau, A., Lee, Y. M., & Merat, N. (2026). The impact of N-back-induced mental workload and time budget on takeover performance. *Accident Analysis & Prevention*, 225, 108327.
- International Organization for Standardization. (2016). *Road vehicles—Transport information and control systems—Detection-response task (DRT) for assessing attentional effects of cognitive load in driving (ISO Standard No.17488:2016).*
- Klauer, S. G., Han, S., Guo, F., & Fincannon, T. (2023). Exploring Driver Adaptation to Lower Levels of Automation (L2) Using Existing Naturalistic Driving Data. *27th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration*, (23–0322).
- Louw, T., Markkula, G., Boer, E., Madigan, R., Carsten, O., & Merat, N. (2017). Coming back into the loop: Drivers' perceptual-motor performance in critical events after automated driving. *Accident Analysis & Prevention*, 108, 9–18.
- Mueller, A. S., Cicchino, J. B., & Calvanelli, J. V. (2024). Habits, attitudes, and expectations of regular users of partial driving automation systems. *Journal of Safety Research*, 88, 125–134. <https://doi.org/10.1016/j.jsr.2023.10.015>
- SAE International. (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.*



Virtuocity
A centre for city simulation

THANK YOU

[Preprint] Qin, H., Gonçalves, R., Goodridge, C. M., & Merat, N. (2026). Eye-Tracking Indices of Cognitive Load Predict Reaction Time to Urban Traffic Hazards: Machine Learning Classification and Regression Across Manual and Level 2 Assisted Driving.

<https://doi.org/10.2139/ssrn.7240439>

*Presented by **Rafael C. Gonçalves** on behalf of the authors.*

Contact:

Hao Qin, PhD student at Institute for Transport Studies, University of Leeds

tshq@leeds.ac.uk