Cognitive Distraction Detection Using Gaze and Pupil with an Interpretable Approach

Kimimasa Tamura, Simon Stent, John Gideon, Kohei Shintani, Guy Rosman
Toyota Research Institute, Cambridge MA, USA

Data & code available!

Why?

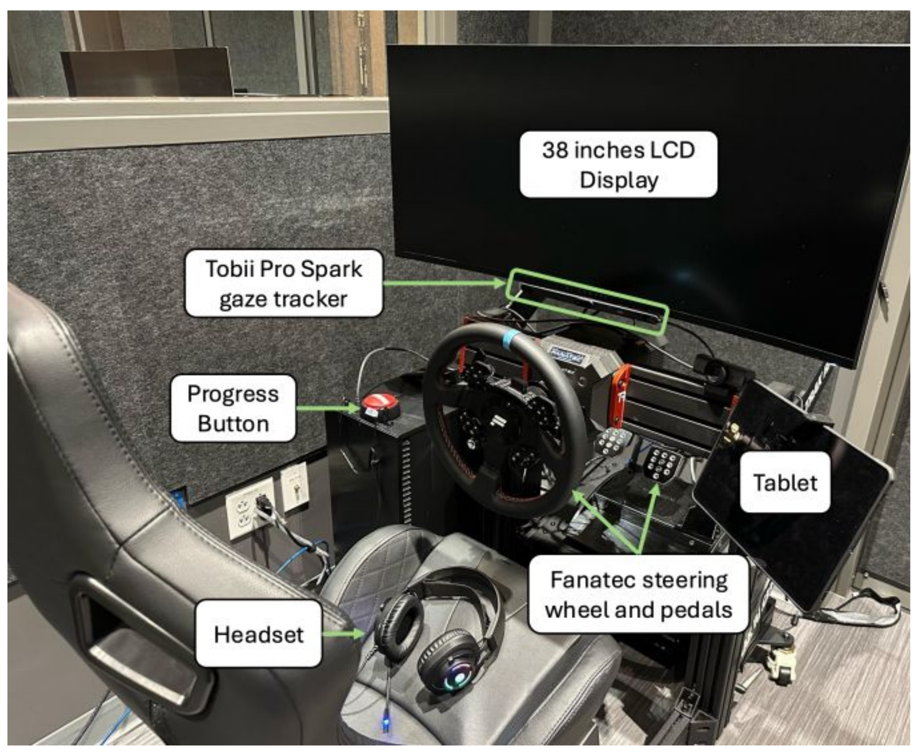
- Distraction causes 8% of fatal crashes US, 2022 (manual, visual, cognitive)
- Cognitive distraction is subtle and lacks standardized detection methods
- In-vehicle deployment requires both **accuracy** and **interpretability**

Goal

Explore **good models** — *for accuracy*

analyze **features** — *for interpretability*

Dataset & Tasks



- 52 participants in driving sim (CARLA Town15)
- 8 urban drives each (≈ 25 min/participant)
- Cognitive Distraction tasks:
  - n-back (1-back digits, 2.5s cadence)
  - statement (plausibility Q&A)
- ~20k (20s) samples in final dataset
- Test using held-out subjects

Raw & Derived Signals

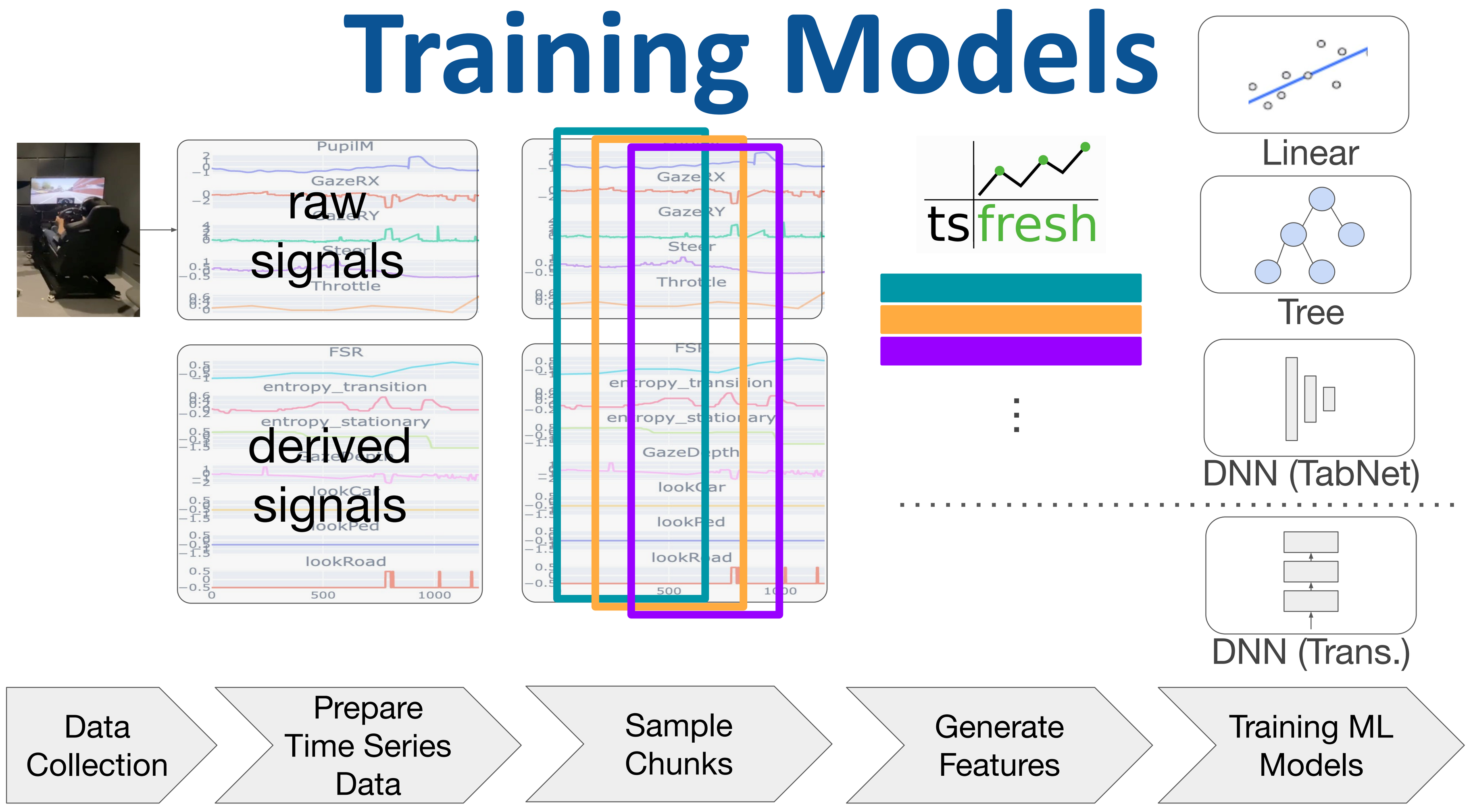
Raw Signals (60 Hz)

- Pupil diameter
- 2D Gaze projection
- Steering angle

Derived Signals (5s sliding windows -> 60Hz)

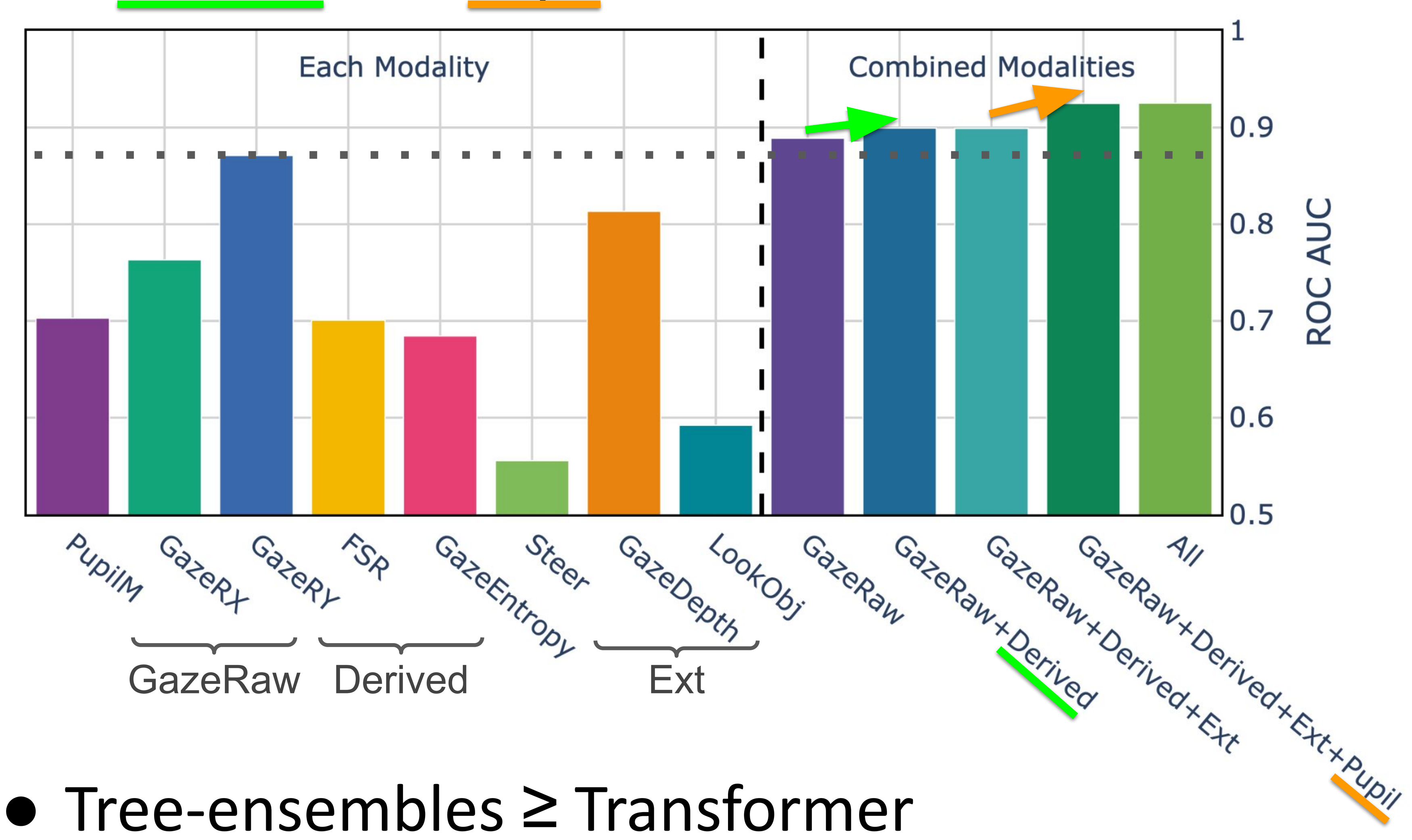
- Fixation–Saccade Ratio
- Gaze Transition / Stationary Entropy
- External: Gaze Depth & Semantic Object Class

Training Models



Key Findings

- Multimodal beats unimodal
  - Derived and Pupil increased AUC with GazeRaw



- Tree-ensembles ≥ Transformer

	LM		GBDT			DNN	
	Log. Reg.	SVM	Light	Xg	Cat	TabNet	Trans.
n-back	0.875	0.775	0.925	0.928	**0.931**	0.892	0.910
state.	0.691	0.599	0.724	0.727	**0.732**	0.724	0.738

- Cross-task generalisation
  - Indicates common CD features across tasks

		Test	
		n-back	statement
Train	n-back	0.931	0.732
	statement	0.914	0.771

Feature Insights

- Non-linearity of vertical gaze DOWN : CD UP
- Baseline pupil diameter UP : CD UP
- Minimum gaze distance UP : CD UP
- Features interact strongly

Takeaways

- Multiple ocular cues + decision-tree ensembles = practical, explainable CD detection
- Model generalises across different cognitive tasks