Detection of Alcohol-Impaired Driving using Eye-Tracking and Vehicle Data via Machine Learning

Ryan Miller[1] Andy Chestovich[1] Anshul Satav[1] Tim Brown[2]

[1]Grinnell College Department of Statistics

[2]University of Iowa Driving Safety Research Institute

Introduction

- Alcohol-involved crashes are a significant problem ($\sim$13,000 fatalities in the US per year, NHTSA)
  - Passive detection of alcohol use may reduce their prevalence
- Using driver monitoring data, we assess:
  - Overall capacity and relative contributions of **eye** and **vehicle** data in differentiating between sober and alcohol-dosed driving samples
  - Measurement duration and performance
  - Robustness when applied to new driving environments

Data Collection

- **Design**: $N = 36$ participants, baseline (sober) drive, dosing to target BrAC of 0.1, four follow up drives at 0.1, 0.085, 0.07, and 0.055 BrAC
- **Data**: time-series of vehicle inputs/states (miniSim) and eye behavior (DMS)
- **Aim**: accurately classify data samples from baseline (alcohol negative) vs. follow up drives (alcohol positive)

Grinnell College
Statistics

IOWA Driving Safety Research Institute

Data Preparation

- "Rural Straight" scenario
  - Approximately 8 min of driving on a straight, two-lane rural road (55 mph)
  - 180 drives, 153 used for analysis
- Drives split into 15, 30, 45, 60, 75, and 90 second samples allowing 50% overlap
  - All samples from a drive assigned to either training (2/3) or validation (1/3) 10x

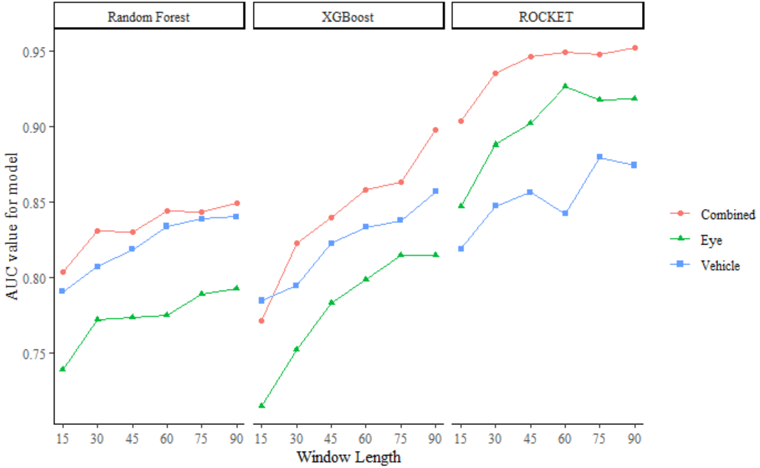
Machine Learning Methods

1. **Feature engineering**
   - Reduce each sample (time-series) to a single data point of derived eye and vehicle measures
   - Model input is a "standard" data matrix (# samples, # derived features)
   - *Random forest* (Ensemble of Decision Trees) and *XGBoost* (Extreme Gradient Boosting)
2. **Deep learning**
   - Uses raw multivariate time-series inputs from each sample
   - Model input is a 3-d array (# samples, # features, # time steps)
   - *ROCKET* Random Convolutional Kernel Transform paired with a *Logistic Regression* final classifier

Feature Engineering

- Inputs chosen based on previous research
  - *Eyes*: pupil size, gaze direction, head position, and blink
  - *Vehicle*: speed, lateral deviation, steering, acceleration, and braking
- Comprehensive summary statistics (dispersion focused) for every input
  - mean, std dev, 5th and 9th percentiles, 90% interpercentile range, skewness, and kurtosis
- Additional derived features
  - PERCLOS for each eye, number and duration of unique fixations, and fixation velocities

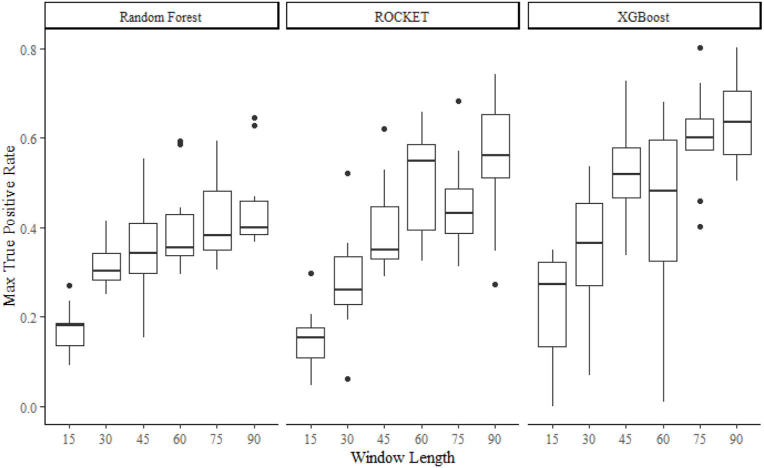
Classification Performance by Duration and Input



AUC-ROC Summary (90 sec)

Predictors	Model	Mean	Std	Min	Max
Eye	Random Forest	0.802	0.061	0.679	0.882
	XGBoost	0.833	0.051	0.741	0.898
	ROCKET	**0.916**	**0.039**	**0.842**	**0.977**
Vehicle	Random Forest	0.848	0.032	0.798	0.883
	XGBoost	0.865	0.039	0.788	0.902
	ROCKET	**0.880**	**0.034**	**0.818**	**0.925**
Combined	Random Forest	0.852	0.062	0.713	0.922
	XGBoost	0.902	0.038	0.830	0.952
	ROCKET	**0.955**	**0.027**	**0.906**	**0.986**

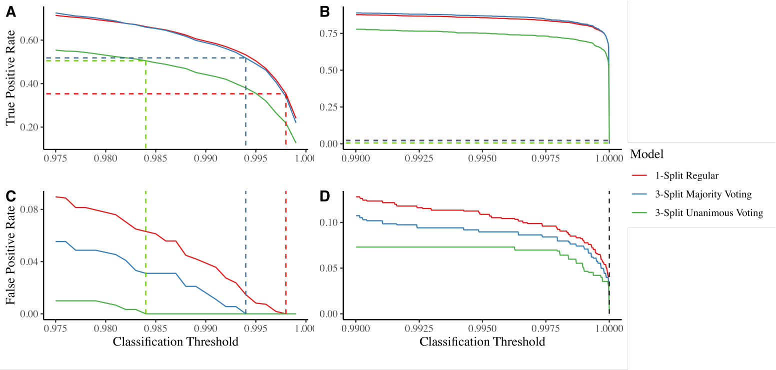
True Positive Rate w/ Zero False Positives



Grinnell College Statistics

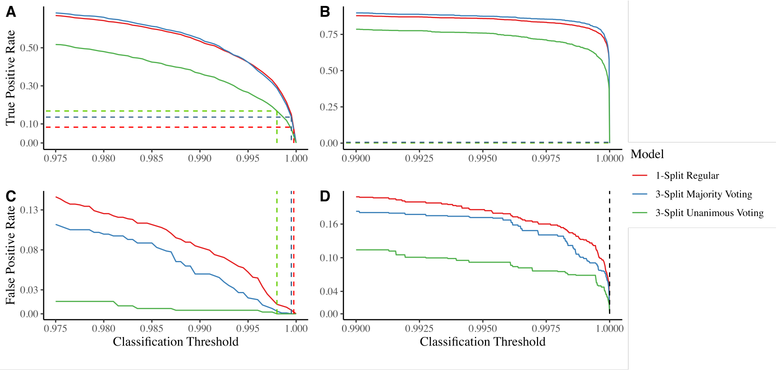
IOWA Driving Safety Research Institute

Voting Schemes and TPR/FPR for Combined Features



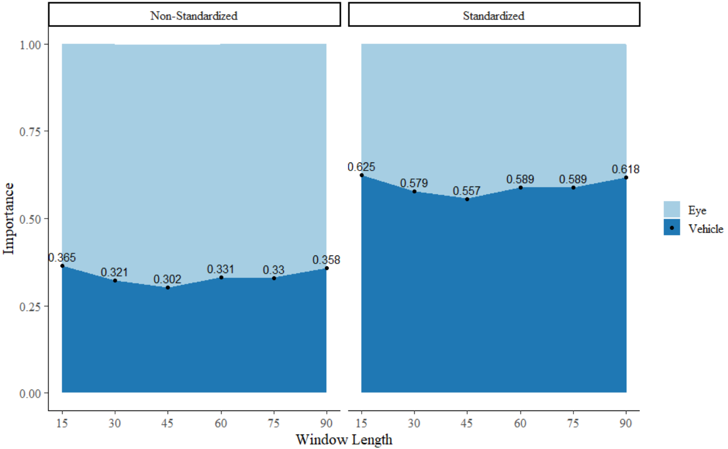
Combined eye and vehicle models (90s), XGBoost (A and C) and ROCKET (B and D)

Voting Schemes and TPR/FPR for Eye Features



Eye-only models (90s), XGBoost (A and C) and ROCKET (B and D)

Feature Contributions (overall eye vs. vehicle)



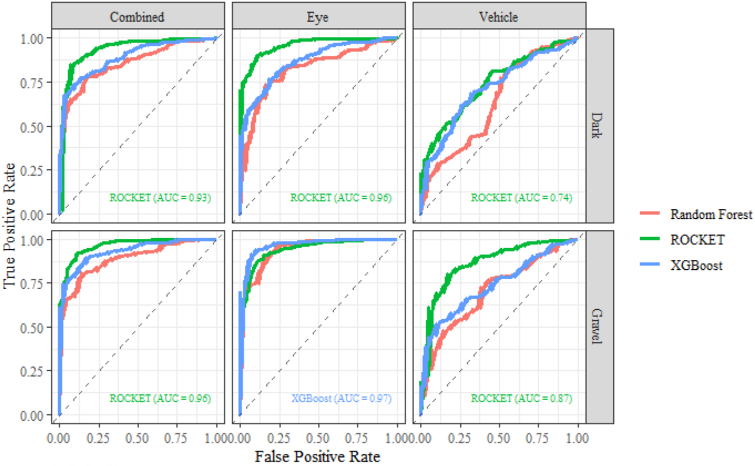
Standardization adjusts for the larger number of eye-inputs available.

Feature Contributions (XGBoost 90s)

Predictor Type	Contribution	Standardized Contribution	n
Lateral Vehicle Control	**0.253**	**0.291**	21
Pupil and Blink Behavior	0.237	0.239	24
Gaze Direction & Eye Movement	0.229	0.092	60
Head Position	0.128	0.147	21
Longitudinal Vehicle Control	0.100	0.127	19
Fixation Behavior	0.047	0.103	11

Generalizability

"Dark" has curves, "Gravel" has curves, a different surface, and no posted speed limit



Summary of Main Results

- Longer measurement duration (90s) lead to better results than shorter ones, even when the total amount of data is fixed (8 minutes)
- Using both eye and vehicle data works best, but eye data is more generalizable to new driving scenarios without model retraining
- Deep learning methods (ROCKET) outperform feature engineering despite modest amounts of data, but handcrafted features and XGBoost are competitive when a very low FPR is desired
- Allowing different intervals within the same drive to "vote" may be useful in reducing false positives

Limitations

- Trained models using study conditions as labels
  - Some "positive" samples had BrAC of 0.055-0.07 with modest drowsiness, others had BrAC of 0.085-0.10 with minimal drowsiness
  - No "negative" samples with low but non-zero BrAC < 0.05
  - Future work should explore a broader range of BrAC and fatigue combinations
- miniSim without ADAS, so possible behavioral differences compared to real, on-road driving
  - Limited number of driving environments (did not consider urban settings or multi-lane highways)

Acknowledgements

- The data used in this research was collected using funding from the Insurance Institute for Highway Safety (IIHS)
- The driver monitoring system and associated technical support were provided by Seeing Machines Ltd.
- The content of this presentation reflects the views of the authors and not necessarily those of IIHS or Seeing Machines.