Collaborative Sciences Center for ROAD SAFETY

www.roadsafety.unc.edu

Safety Enhancement by Detecting Driver Impairment through Analysis of Real-time Volatilities

Collaboration: University of Tennessee, Knoxville

May 16, 2024

R44 Project Team

UT Knoxville

- Dr. Asad Khattak
- Dr. Subhadeep Chakraborty
- Dr. Iman Mahdinia
- Sheikh M. Usman
- + Contributions by several students, especially Dr. Numan Ahmed, Dr. Ramin Arvin, & Mr. Riley Tavassoli

THE UNIVERSITY OF TENNESSEE KNOXVILLE

R44 Project Outcomes

- Usman, S.M., Asad Khattak, Subhadeep Chakraborty, Iman Mahdinia, and Riley Tavassoli (2024). Detection of Distracted Driving through the Analysis of Real-time Driver, Vehicle and Roadway Volatilities. Forthcoming *Journal of Transportation Safety and Security*.
- Ahmad, Numan, Ramin Arvin, and Asad Khattak, (2023). How is the duration of distraction related to safety–critical events? Harnessing naturalistic driving data to explore the role of driving instability. *Journal of Safety Research*, 27.
- Ahmad, Numan, Ramin Arvin, and Asad J Khattak (2023). Exploring pathways from driving errors and violations to crashes: the role of instability in driving. *Accident Analysis & Prevention*, 179, 106876.
- Ahmad Numan, Asad Khattak, and Hamparsum Bozdogan (2023). Predicting safety-critical events using driver behaviors and performance: Application of machine learning. Transportation Research Board, Paper TRBAM-23-00144.
- Ahmad Numan, "Role of Human Factors, Driving Instability, and Roadway Environment in Safety Critical Events: Safe System Approach." Ph.D. dissertation, The University of Tennessee, 2021. https://trace.tennessee.edu/utk_graddiss/6961

Overview: Research Objectives

The project aims to understand detection of driver impairment using streaming biometric information

The key project objectives are:

- **Collect unique high-frequency multi-dimensional large-scale data** using sensors that monitor the driver, vehicle, and roadways.
- Harness the data
  - Quantify variations in driver biometrics and behavior, vehicle kinematics, & roadway/env. conditions
  - Utilize the concept of volatility as a leading indicator of crash risk
  - Analyze correlations of driver biometrics and driving style with driver impairment and crash risk
- Develop algorithms to identify driving impairment by monitoring data streams emanating from the driver, vehicle, and roadway in real-time to provide feedback and warnings to drivers & surrounding vehicles

Overview: Research Questions

The research questions related to distracted driving events are:

- Can driver distraction be identified using biometric and vehicle-based sensors in different driving scenarios?
- How can driving events be classified as normal and distracted/impaired based on volatility measures in data streams?
- How is prolonged distracted driving associated with driving instability and safety-critical events?
- What are the mechanisms for driving errors and violations that lead to safety-critical events?

R44 Project: Studies Conducted

- Study I:
  Detection of Distracted Driving through the Analysis of Real-time Driver, Vehicle, and Roadway Volatilities.

- Study II:
  How is the Duration of Distraction-related to safety-critical events? Harnessing naturalistic driving data to explore the role of driving instability.

- Study III:
- Exploring Pathways from Driving Errors and Violations to Crashes: The Role of Instability in Driving.

- Study IV:
- Predicting Safety-Critical Events using Driver Behaviors and Performance: Application of Machine Learning.

Study 1 (Project R44)

Detection of Distracted Driving through the Analysis of Real-time Driver, Vehicle, and Roadway Volatilities

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Collaborative Sciences Center for ROAD SAFETY

Introduction

Types of Driving Distraction

- Visual
- Cognitive
- Auditory
- Physical

- Detection Response Task (DRT) used to assess the attentional effect of secondary tasks on driving performance

- Driver response to visual or tactile stimuli presented to drivers at random intervals provide a measurable indicator of distraction

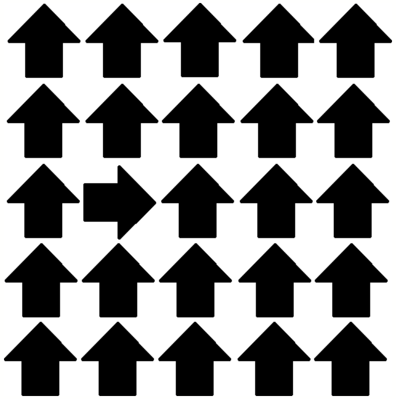
Study Objective

To detect events of distracted driving under normal and different distracted driving

- Driving simulator was used
- Scenarios were developed using the visual DRT (grid of arrows) with varying difficulty levels

Driver’s view inside the vehicle

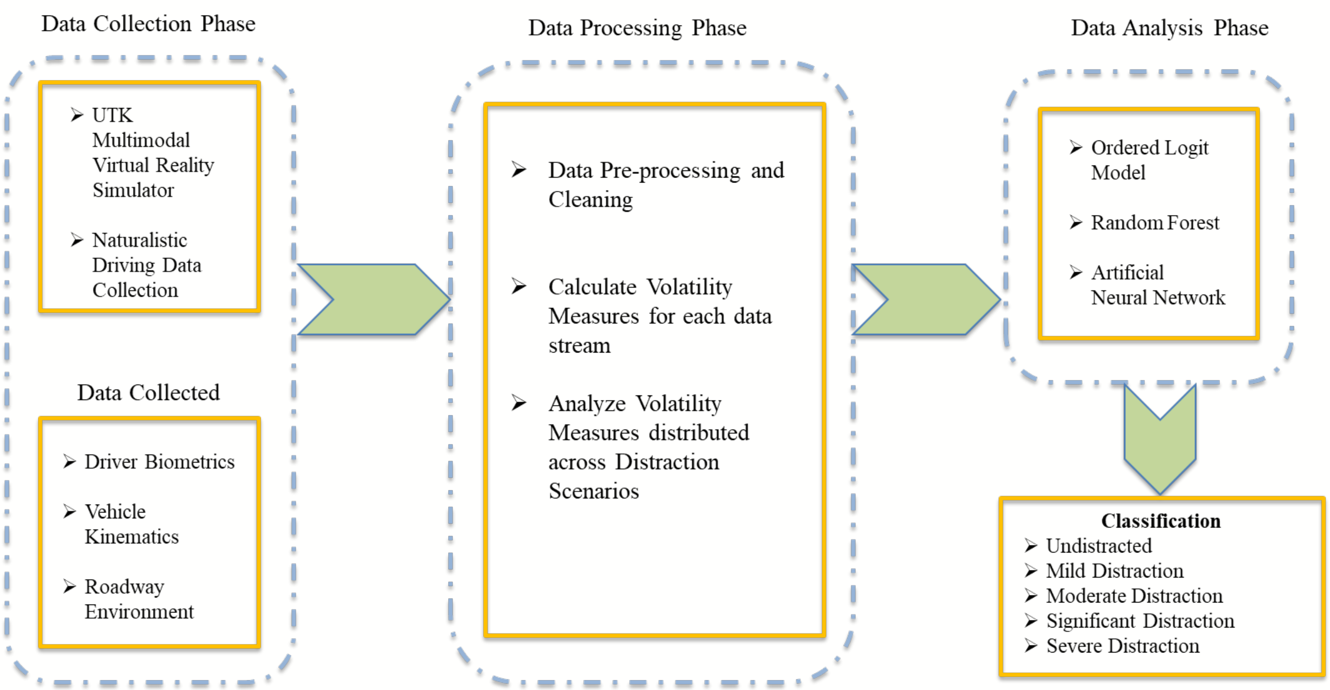
Driving simulator hardware



Grid of Arrows

Study Framework

- Driving Simulation done on Multimodal Virtual Reality Simulator
- Captured Driver Biometrics, Vehicle Kinematics, and Roadway Surroundings
- 4x4, 5x5, 6x6, 8x8 grids of arrows used as distraction scenarios



Overall study framework

Methodological Framework

- Data Collected (N=617)-Dependent Variable
  - Level of Distraction (Undistracted, Mild, Moderate, Significant, Severe Distraction)
- Estimated Panel Ordered Logit Model
  - Ordered nature of the response variable
  - Repeated observations over time for the same subjects
- Applied Machine Learning-Random Forest
  - Captures complex non-linear behavior
  - Prediction of Distraction is important
- Applied Artificial Neural Network
  - Identification of intricate patterns in multidimensional data
  - Ability to learn from raw data with minimal preprocessing

Descriptive Statistics

Means of volatility measures in driver biometrics, vehicle kinematics, and roadway surroundings increase with higher levels of distraction

Volatility Measures	Undistracted N = 204		Mild Distraction N= 123		Moderate Distraction N = 120		Significant Distraction N = 85		Severe Distraction N = 85	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CV-Eye Movement	0.40	0.330	0.68	0.333	0.79	0.29	0.83	0.28	0.94	0.27
CV-Speed	0.06	0.05	0.08	0.06	0.10	0.068	0.11	0.07	0.12	0.11
MAD-Acceleration	1.47	1.44	1.56	1.53	1.63	1.55	1.86	1.73	1.95	2.00
CV-Centre Distance	0.15	0.14	0.20	0.18	0.34	0.19	0.35	0.21	0.38	0.23
CV-Front Distance	0.08	0.10	0.09	0.105	0.094	0.106	0.11	0.11	0.13	0.13
CV-Back Distance	0.04	0.07	0.056	0.08	0.059	0.10	0.07	0.15	0.10	0.15

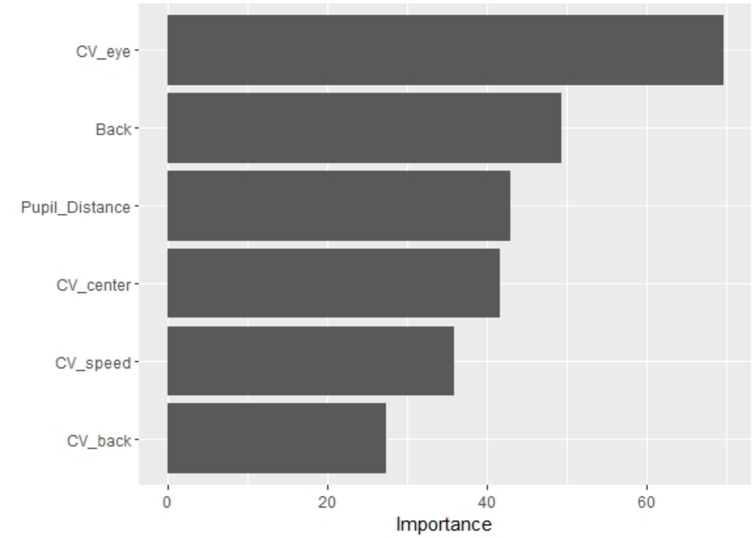
Study Results

- Ordered Logit Model
- Training Data (N = 507)
- Volatility indicators found statistically significant

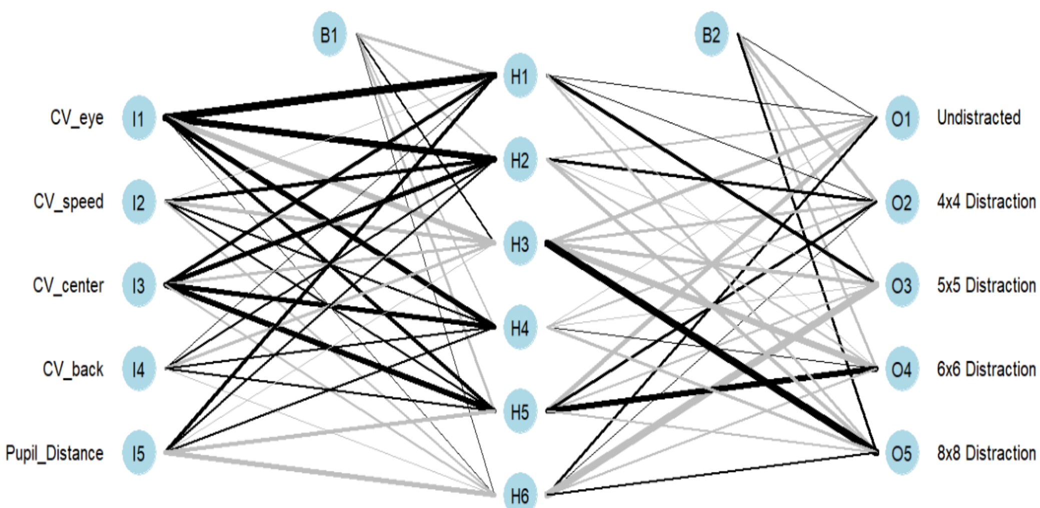
Variables	Coefficient	t-stat	p-value	Marginal Effects (Level of Distraction)				
				Undistracted	Mild	Moderate	Significant	Severe
CV-Eye Movement	6.428	10.68	0.000	-0.8865	-0.0612	0.1886	0.2816	0.4775
CV-Speed	2.223	4.08	0.000	-0.3065	-0.0211	0.0652	0.0974	0.1651
CV-Centre Distance	4.051	6.99	0.000	-0.5586	-0.0385	0.1188	0.1774	0.3009
CV-Back Distance	1.747	2.04	0.041	-0.2409	-0.0166	0.0512	0.0765	0.1297
Eye Movement	1.142	2.05	0.041	-0.1574	-0.0108	0.0335	0.0500	0.0848
Thresholds								
μ_1	-0.983							
μ_2	0.0668							
μ_3	0.9734							
μ_4	2.007							
Summary Statistics								
N	507							
LL at convergence	-627.4721							
LL at null	-770.3974							
Pseudo R^2	0.185							
$\chi^2(5)$	285.85							
Prob> $\chi^2(5)$	0.000							
AIC	1274.944							
BIC	1317.229							

Study Results

- High importance: Coefficients of Variation in Eye Movements; Vehicle Distance from Centerline; & the following vehicle
- Random Forest: Highest Prediction Accuracy of 77.27%



Variable Importance Plot -Random Forest



Artificial Neural Network

Random Forest

Observed Outcomes	**Predicted Outcomes** Undistracted	Mild Distraction	Moderate Distraction	Significant Distraction	Severe Distraction	**Total**
Undistracted	29	5	2	1	1	**38**
Mild	2	11	1	0	1	**15**
Moderate	0	4	19	2	1	**26**
Significant	0	2	3	10	0	**15**
Severe	0	0	0	0	16	**16**
Total	**31**	**22**	**25**	**13**	**19**	
Performance Metrics						
Accuracy			77.27%			
Precision	0.9354	0.5000	0.7600	0.7692	0.8421	
Recall	0.7631	0.7333	0.7307	0.6667	1.0000	
F1 Score	0.8405	0.5946	0.7451	0.7142	0.9143	

Confusion Matrix- Random Forest

Future Research

- Variations in driver biometrics (driver gaze, eye openness), vehicle kinematics, & surrounding → leading distraction indicators
- Findings emphasize
  - Development of proactive safety measures
  - Driver feedback systems – safety warnings, control assists, & automation
- Development of algorithms-promising results of Random Forest classifier in vehicles to detect distracted driving

Study II (Project R44)

How is the Duration of Distraction-related to Safety-Critical Events?

Harnessing Naturalistic Driving Data to explore the role of Driving Instability

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Collaborative Sciences Center for ROAD SAFETY

Introduction

- Distracted driving is a critical safety concern
- At 50 mph, sending/reading a text for 5 seconds is equivalent to driving the length of a football field (360 ft) with eyes closed
- Fundamental understanding of how distractions lead to crashes is needed to develop appropriate countermeasure strategies

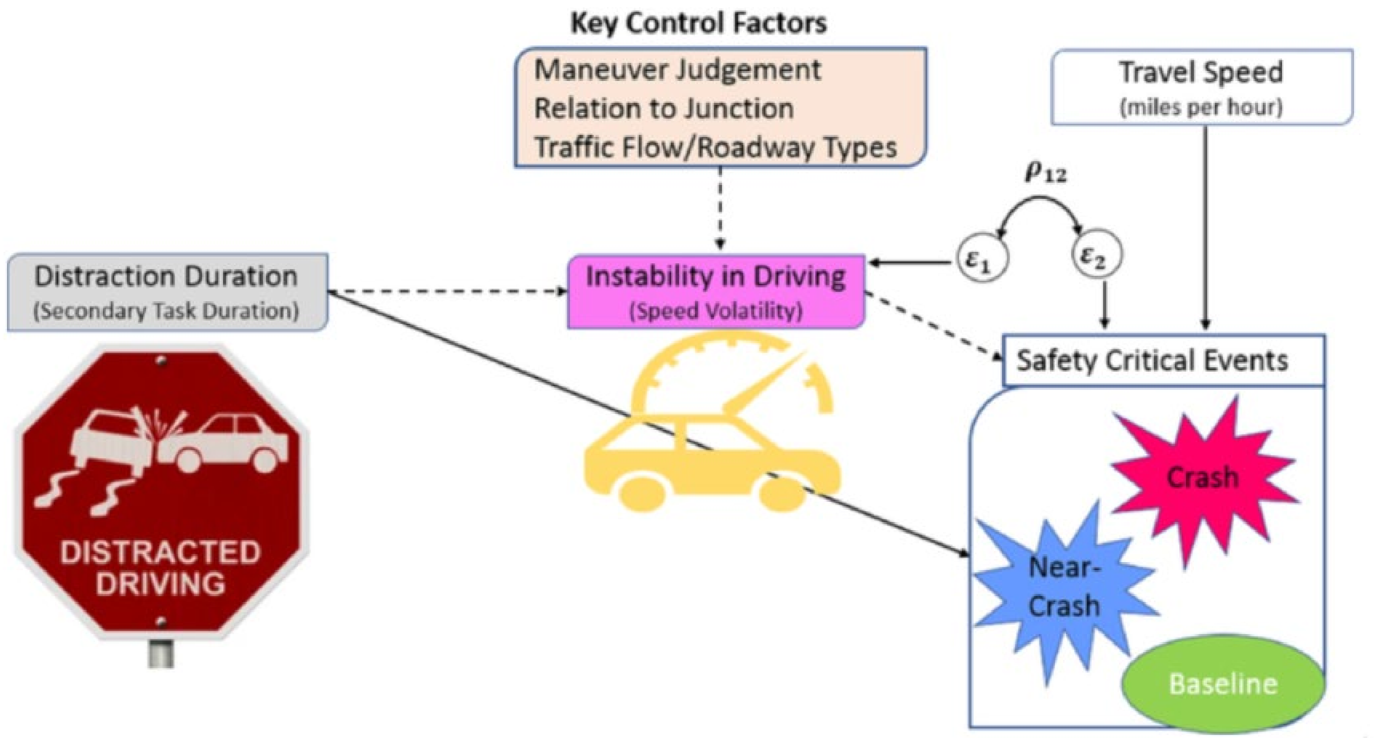
Data Used

- The study analyzed a subsample of the SHRP-2 naturalistic driving data
  - Provides real-world data on pre-crash driving behaviors
  - Secondary tasks (along with their duration)
  - Vehicle kinematics
  - Roadway and environment variables

SHRP2
STRATEGIC HIGHWAY RESEARCH PROGRAM

Objectives and Methodology

- To understand how distraction duration relates to SCEs
- To capture any non-linear effects of distraction duration
- Path Analysis applied to safety data
- Direct & Indirect Effects of distracted driving captured in the study



Study Conceptual Framework

Modeling results: Joint estimation

Independent Variables	Tobit Model (COV of Speed)					Ordered Probit Model (Event Outcome)				
	Coeff.	t-stat	ME-1	ME-2	ME-3	Coeff.	t-stat	MEs		
								Baseline	Near-Crash	Crash
COV of speed	---	---	---	---	---	3.8125	32.14	-0.7416	0.4016	0.3400
Secondary Task Duration (seconds)	-0.0071	-3.08	-0.0055	-0.0072	-0.0041	-0.0368	-2.88	0.0072	-0.0039	-0.0033
Secondary Task Duration * Secondary Task Duration	0.0029	11.01	0.0022	0.0029	0.0017	0.0087	5.72	-0.0017	0.0009	0.0008
Speed (miles per hour)	---	---	---	---	---	0.0495	6.33	-0.0096	0.0052	0.0044
Speed (miles per hour) * Speed (miles per hour)	---	---	---	---	---	-0.0012	-5.94	0.0002	-0.0001	-0.0001
Relation to Junction ***(Base Category = Intersection or intersection related)***										
Other (rail grade crossing, parking lot entrance or exit etc.)	-0.0329	-2.86	-0.0287	-0.0248	-0.0224	---	---	---	---	---
Parking lot, within boundary	-0.0177	-0.71[b]	-0.0156	-0.0130	-0.0122	---	---	---	---	---
Driveway, alley access, etc.	-0.1324	-10.91	-0.1097	-0.1203	-0.0831	---	---	---	---	---
Entrance/Exit ramp or interchange area	-0.1601	-12.93	-0.1305	-0.1522	-0.0982	---	---	---	---	---
Non-junction	-0.1565	-20.47	-0.1279	-0.1479	-0.0963	---	---	---	---	---
Maneuver Judgement ***(Base Category = Unsafe and illegal)***										
Unsafe but legal	0.0254	1.33[b]	0.0228	0.0162	0.0184	---	---	---	---	---
Safe but illegal	-0.1108	-5.10	-0.0936	-0.0925	-0.0725	---	---	---	---	---
Safe and legal	-0.1332	-11.58	-0.1112	-0.1155	-0.0856	---	---	---	---	---
Traffic flow ***(Base Category = Divided (median Strip or barrier)***										
No lanes	0.2210	9.13	0.1876	0.1796	0.1461	---	---	---	---	---
Not divided - center 2-way left turn lane	0.0241	2.37	0.0183	0.0265	0.0135	---	---	---	---	---
Not divided - simple 2-way trafficway	0.0537	8.36	0.0417	0.0569	0.0309	---	---	---	---	---
One-way traffic	0.1179	8.26	0.0952	0.1137	0.0717	---	---	---	---	---
Road Surface Condition: Wet (1/0)	0.0242	3.04	0.0190	0.0248	0.0142	0.1083	2.58	-0.0211	0.0114	0.0097
Constant	0.3956	30.55	---	---	---	---	---	---	---	---
Sigma (e. Coefficient of Variation of Speed)	0.2688	134.82	---	---	---	---	---	---	---	---
Thresholds										
μ_1	---	---	---	---	---	2.2306	27.58	---	---	---
μ_2	---	---	---	---	---	3.1284	37.27	---	---	---
Models Summary										
Rho (Correlation between residuals of the two equations)	-0.4698 (t-stats = -12.18)									
AIC	10375.73									
BIC	10561.14									

Note: * and ** indicate that a particular variable showed partial statistical significance and insignificance, respectively.
ME-1 which refers to both censored and uncensored observations;
ME-2 indicates probability of being uncensored; ME-3 is similar to ME-1 but based on only uncensored observations.

Path analysis results using predictive margins

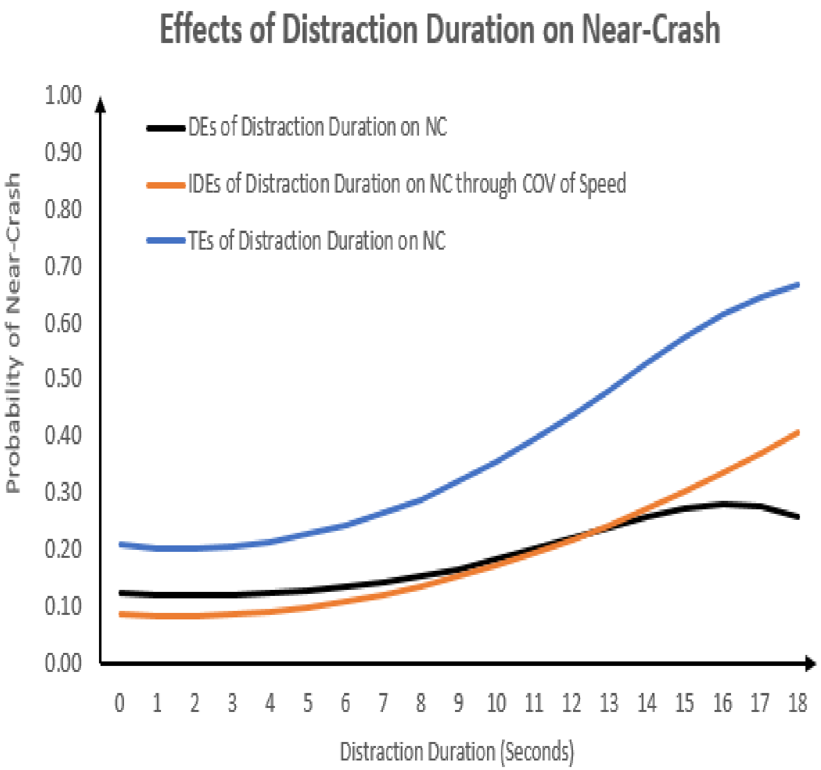
Distraction duration versus Probability of Crash

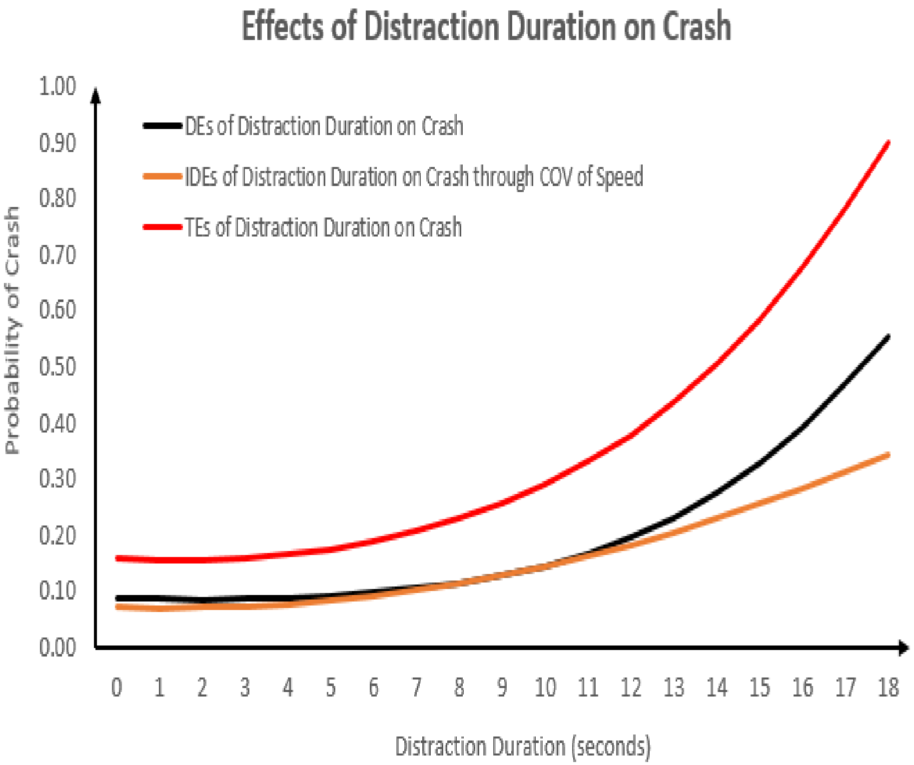
Secondary Task Duration (seconds)	Direct Effects of Secondary Task Duration on Crash		Effects of Secondary Task Duration on COV of Speed		Indirect Effects of Secondary Task Duration on Crash through COV of Speed	Total Effects of Secondary Task Duration on Crash
	Predictive MEs	t-stats	Predictive MEs	t-stats		
0	0.0889	15.83	0.2117	56.53	0.0720	0.1609
1	0.0865	15.57	0.2075	67.73	0.0706	0.1571
2	0.0856	15.37	0.2091	62.67	0.0711	0.1567
3	0.0862	15.42	0.2163	55.33	0.0736	0.1597
4	0.0882	15.85	0.2294	51.88	0.0780	0.1662
5	0.0919	16.78	0.2481	51.51	0.0844	0.1763
6	0.0974	18.33	0.2726	52.29	0.0927	0.1901
7	0.1049	20.54	0.3029	51.88	0.1030	0.2079
8	0.1151	22.86	0.3389	48.88	0.1152	0.2303
9	0.1283	23.32	0.3806	43.95	0.1294	0.2577
10	0.1455	20.33	0.4280	38.71	0.1455	0.2910
11	0.1675	15.97	0.4812	34.11	0.1636	0.3311
12	0.1955	12.37	0.5402	30.39	0.1837	0.3791
13	0.2308	9.86	0.6049	27.45	0.2057	0.4365
14	0.2750	8.18	0.6753	25.14	0.2296	0.5046
15	0.3292	7.07	0.7514	23.30	0.2555	0.5847
16	0.3943	6.37	0.8333	21.83	0.2833	0.6777
17	0.4702	6.00	0.9210	20.63	0.3131	0.7833
18	0.5549	5.92	1.0144	19.64	0.3449	0.8998

Distraction duration versus Probability of Near-crash

Secondary Task Duration (seconds)	Direct Effects of Secondary Task Duration on Near-Crash		Effects of Secondary Task Duration on COV of Speed		Indirect Effects of Secondary Task on Near-Crash through COV of Speed	Total Effects of Secondary Task on Near-Crash
	Predictive MEs	t-stats	Predictive MEs	t-stats		
0	0.1230	36.58	0.2117	56.53	0.0850	0.2080
1	0.1200	39.98	0.2075	67.73	0.0833	0.2033
2	0.1188	38.85	0.2091	62.67	0.0840	0.2028
3	0.1196	36.24	0.2163	55.33	0.0869	0.2064
4	0.1221	33.91	0.2294	51.88	0.0921	0.2143
5	0.1267	31.88	0.2481	51.51	0.0996	0.2263
6	0.1332	29.55	0.2726	52.29	0.1095	0.2427
7	0.1420	26.48	0.3029	51.88	0.1216	0.2636
8	0.1531	22.92	0.3389	48.88	0.1361	0.2891
9	0.1665	19.52	0.3806	43.95	0.1528	0.3194
10	0.1824	16.74	0.4280	38.71	0.1719	0.3543
11	0.2004	14.74	0.4812	34.11	0.1933	0.3937
12	0.2199	13.53	0.5402	30.39	0.2169	0.4368
13	0.2398	13.16	0.6049	27.45	0.2429	0.4827
14	0.2582	13.92	0.6753	25.14	0.2712	0.5294
15	0.2726	16.87	0.7514	23.30	0.3018	0.5743
16	0.2796	25.82	0.8333	21.83	0.3347	0.6142
17	0.2759	27.07	0.9210	20.63	0.3699	0.6457
18	0.2590	11.05	1.0144	19.64	0.4074	0.6664

Path analysis results using predictive margins





TEs refer to total Effects
IDE refer to indirect effects
DEs refer to direct effects

Key Findings

- Instability in driving was higher in crash and near-crash events than in baselines.
- The Coefficient of Variation of speed was higher in crashes and near-crashes than in baselines.
- Distraction duration was longer in crashes than near-crashes and baselines.
- The probability of a crash increases exponentially when the distraction duration exceeds 8 seconds.

Future Research

- Hands-free technologies (voice-activated controls, and virtual assistants) can reduce distracted driving.
- Fixed and dynamic message signs about distracted driving have the potential to reduce both distracted driving and driving instability.
- Deploying multiple vehicle technologies (e.g., forward-collision warning system and adaptive cruise control) can be help reduced driving instability and safety critical events.

Study III (Project R44)

Exploring Pathways from Driving Errors and Violations to Crashes: the Role of Instability in Driving

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Collaborative Sciences Center for ROAD SAFETY

Introduction

- Driver errors and violations substantially contribute to roadway crashes
- The study explores how driving errors, violations, and roadway environments impact of instability in driving speed and safety-critical events.

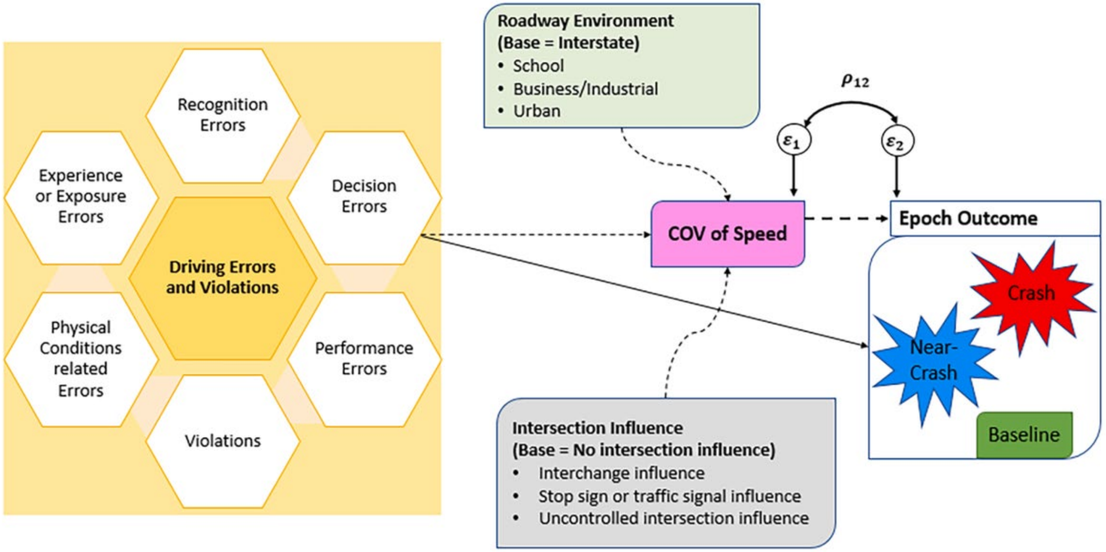
Data and Statistical Techniques Used

- The study analyzed a subsample of the SHRP-2 naturalistic driving study data (N=9,239)
- Analytical techniques:
  - Path analysis
  - Tobit and Ordered Probit regressions to jointly model outcomes

SHRP2
STRATEGIC HIGHWAY RESEARCH PROGRAM

Study Objective

- To identify pathways from driving errors and violations in diverse roadway environments to SCEs through instability in driving speed
- Jointly estimate models that account for potential correlation between the unobserved factors associated with
  - Epoch outcomes (baseline, near-crash, and crash) and
  - Coefficient of variation “COV” of speed (instability in driving speed)



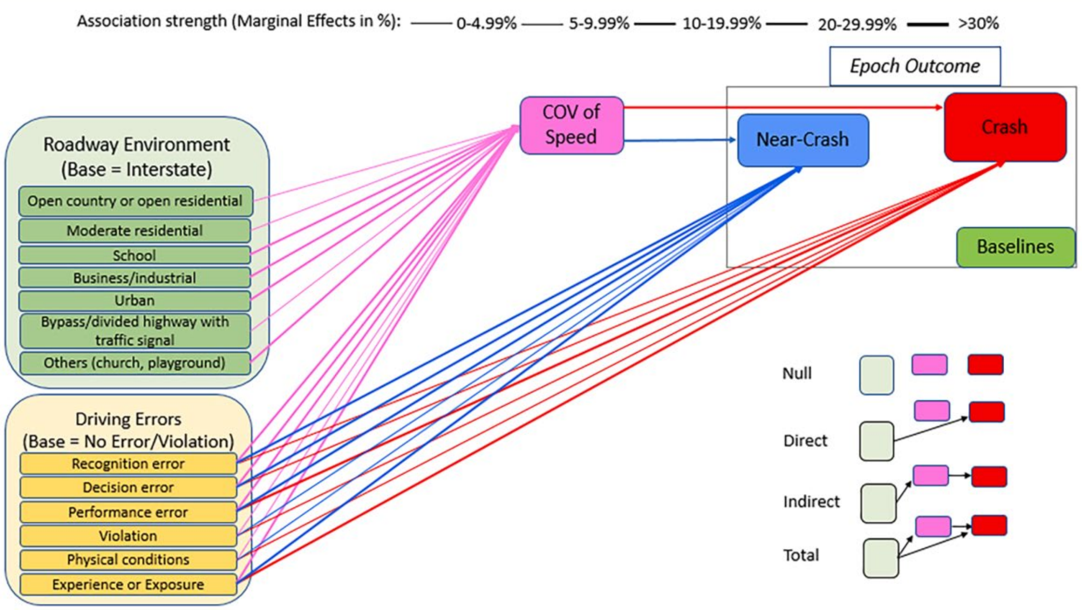
Study Conceptual Framework

Methodology

- Path Analysis

Model 1 (Tobit model): $Driving\ instability$: $Y_1 = \beta_0 + \beta_1 X_1 + \varepsilon_1$
Model 2 (Ordered probit) : $Event\ outcome$: $Y_2 = \beta_0 + \beta_2 X_2 + \gamma Y_1 + \varepsilon_2$

- Driving errors and violations → instability in driving speed (i.e., higher COV of speed) → SCEs.



Path Analysis

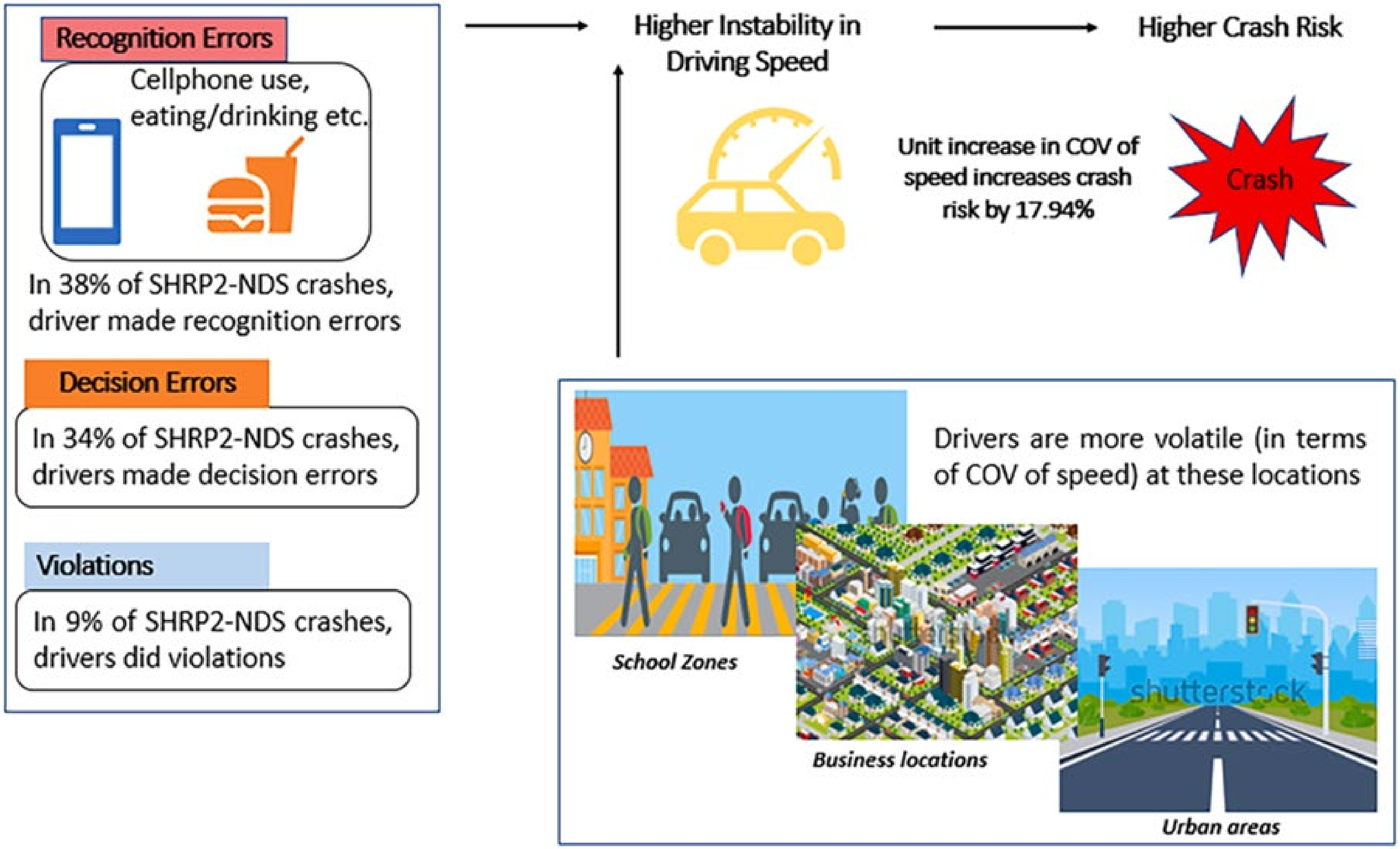
Modeling results: Joint Estimation

Independent variables	Joint Estimation									
	Tobit Model (COV of Speed)					Ordered Probit Model (Event Outcome)				
	Coeff.	t-stat	ME-1	ME-2	ME-3	Coeff.	t-stat	MEs: Baseline	MEs: Near-Crash	MEs: Crash
COV of Speed	---	---	---	---	---	3.0065	29.14	-0.4744	0.2950	0.1794
Drivers Errors (Base Outcome = No Driving Errors)										
Recognition Errors	0.3375	32.50	0.2636	37.16%	0.2064	1.1357	12.60	-0.1792	0.1114	0.0678
Decision Errors	0.1998	19.60	0.1560	22.00%	0.1222	1.3889	19.06	-0.2191	0.1363	0.0829
Performance Errors	0.3454	11.63	0.2697	38.03%	0.2113	1.9086	10.34	-0.3011	0.1873	0.1139
Violations	0.0647	6.57	0.0505	7.12%	0.0396	0.8082	13.43	-0.1275	0.0793	0.0482
Physical Conditions	0.0391*	1.83*	0.0306*	4.31%*	0.0239*	0.7659	6.02	-0.1208	0.0751	0.0457
Experience or Exposure	0.2296	5.04	0.1793	25.28%	0.1404	1.7344	7.36	-0.2737	0.1702	0.1035
Roadway Locality (Base = Interstate)										
Open Country or Open Residential	0.0662	6.86	0.0517	7.29%	0.0405	---	---	---	---	---
Moderate Residential	0.1154	15.16	0.0901	12.71%	0.0706	---	---	---	---	---
School	0.1384	11.73	0.1081	15.24%	0.0847	---	---	---	---	---
Business/Industrial	0.1417	20.15	0.1107	15.60%	0.0867	---	---	---	---	---
Urban	0.2376	15.39	0.1856	26.17%	0.1454	---	---	---	---	---
Bypass or Divided Highway with traffic signals	0.0721	4.96	0.0563	7.94%	0.0441	---	---	---	---	---
Others (e.g., church, playground, and Campground)	0.1573	8.86	0.1228	17.32%	0.0962	---	---	---	---	---
Intersection Influence (Base = No intersection influence)										
Interchange influence	0.1143	8.10	0.0893	12.59%	0.0699	---	---	---	---	---
Stop sign or traffic signal influence	0.1907	26.57	0.1490	21.00%	0.1167	---	---	---	---	---
Uncontrolled intersection influence	0.1646	12.47	0.1285	18.13%	0.1007	---	---	---	---	---
Parking lot or driveways influence	0.2434	20.84	0.1901	26.80%	0.1489	---	---	---	---	---
Others (e.g., crosswalk, railroad crossing, roundabouts)	0.1855	8.44	0.1449	20.43%	0.1135	---	---	---	---	---
Secondary task duration	0.0015*	1.56*	0.0012*	0.16%*	0.0009*	0.0258	4.40	-0.0041	0.0025	0.0015
Level of Service (Base Category = C to F)										
LOS A: Free flow traffic condition	-0.1131	-13.41	-0.0883	-12.45%	-0.0692	---	---	---	---	---
LOS B: Traffic Flow with some restriction	-0.1094	-11.81	-0.0855	-12.05%	-0.0669	---	---	---	---	---
Constant	0.1354	15.06	---	---	---	---	---	---	---	---
Sigma (e. Coefficient of Variation of Speed)	0.2312	135.66	---	---	---	---	---	---	---	---
Thresholds										
μ_1	---	---	---	---	---	2.0769	69.51	---	---	---
μ_2	---	---	---	---	---	3.3533	67.88	---	---	---
Models Summary										
ρ (Rho)	-0.3852 (t-stats = -12.82)									
Number of observations	9,239									
AIC	5875.5970									

All driving errors and violations increase driving instability which in turn increases crash risk

Note: * and ** indicate that a particular variable showed partial statistical significance and insignificance, respectively.
ME-1 which refers to both censored and uncensored observations;
ME-2 indicates probability of being uncensored; ME-3 is similar to ME-1 but based on only uncensored observations.

Key Findings-Graphic

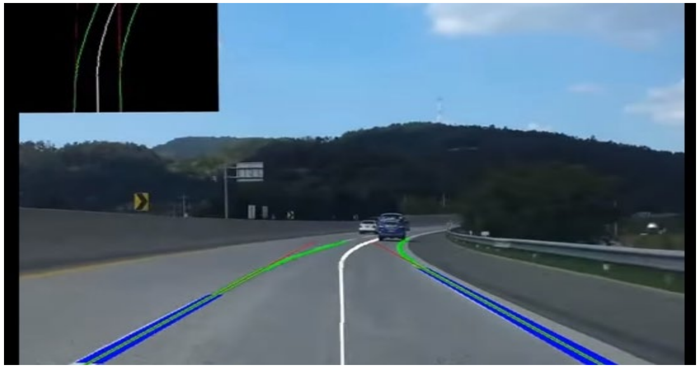


Key Findings

- All five driving error types, and violations were associated with higher instability in driving speed (COV of speed).
- **Performance errors** exhibit the strongest positive correlation with crash risk, followed by **experience errors**, decision errors, and recognition errors.
- Instability in driving speed is significantly higher in urban areas, business/industrial locations, and school zones compared to driving on interstates.
- All driving errors and violations not only contribute to SCEs directly but also indirectly through instability in driving speed.

Future research

- Forward-collision warning systems, adaptive cruise control, lane tracking systems, and lateral vehicle detection system can reduce one or more driving errors.
- Dilemma zone mitigation systems have the potential to reduce a significant percentage of violations.
- Awareness campaigns and mandatory training programs for drivers can reduce performance errors and experience errors.



Study IV (Project R44)

Predicting Safety-Critical Events using Driver Behaviors and Performance: Application of Machine Learning.

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Collaborative Sciences Center for ROAD SAFETY

Introduction

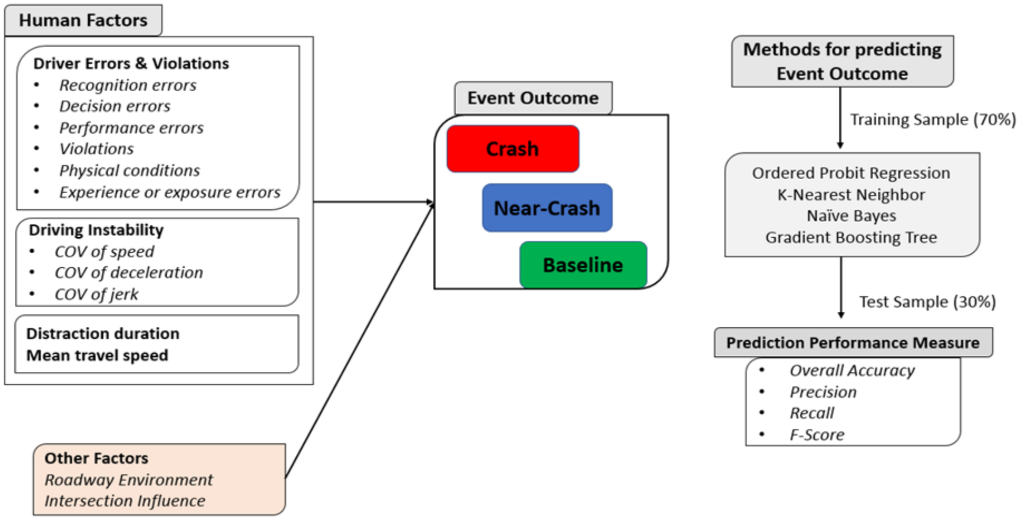
- Human factors give rise to more than 90% of road traffic crashes
- With recent advancements in machine learning (ML) and the availability of detailed Naturalistic Driving Study (NDS) data collected through the SHRP-2 program, new avenues for predicting SCEs can be explored
- Focus on real-time prediction of SCEs using driving errors and violations, distraction duration, driving instability.
- **Data Used**
- The study analyzed a subsample of the SHRP-2 NDS data.
- Sample Size = 9,237 observations

Study Objective

- To enhance predictive accuracy by leveraging the unique SHRP2-NDS data
  - Data provides dynamic pre-crash information on driving behavior and performance

- To deepen the understanding of the connection between pre-crash driving behavior, performance, and SCEs.

- To identify the most accurate model or method for real-time prediction of SCEs.



Study Framework

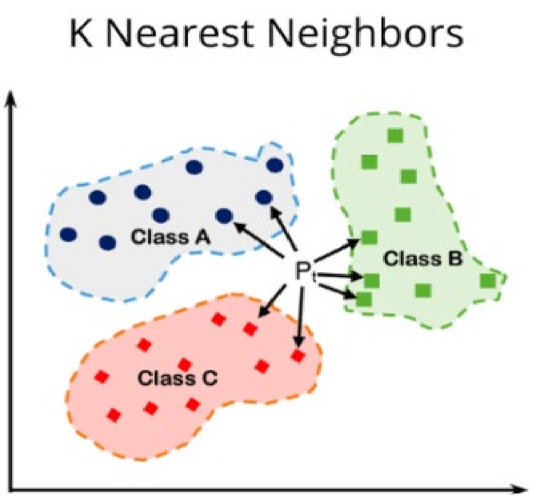
Methodology

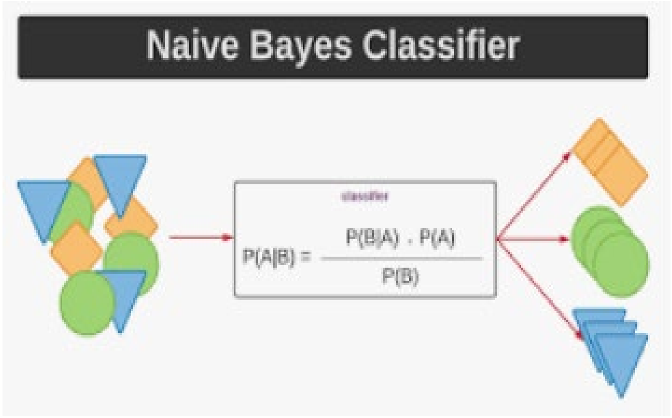
Conventional Statistical Techniques

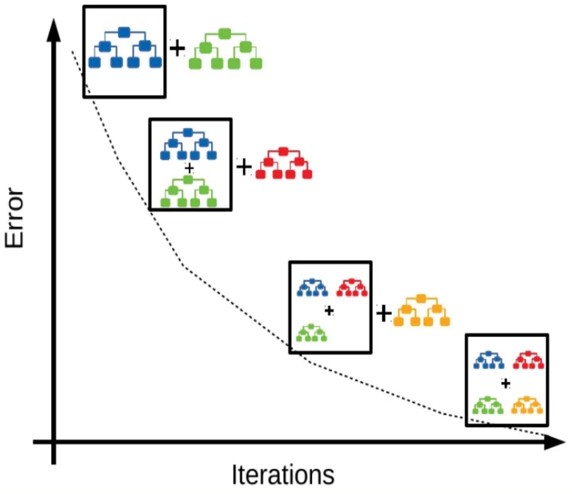
- Ordered Probit Regression

Machine Learning Methods

- K-Nearest Neighbors
- Naïve Bayes
- Gradient Boosting Decision Tree





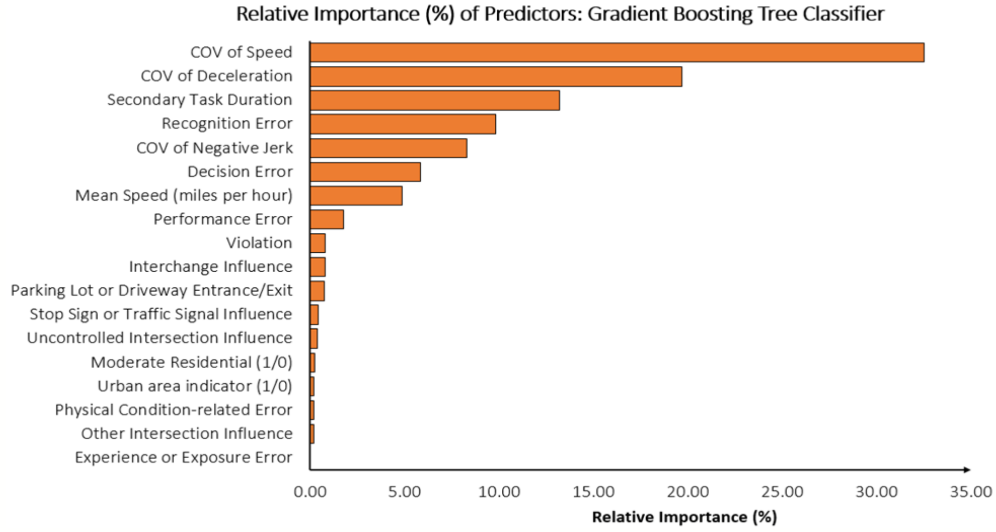


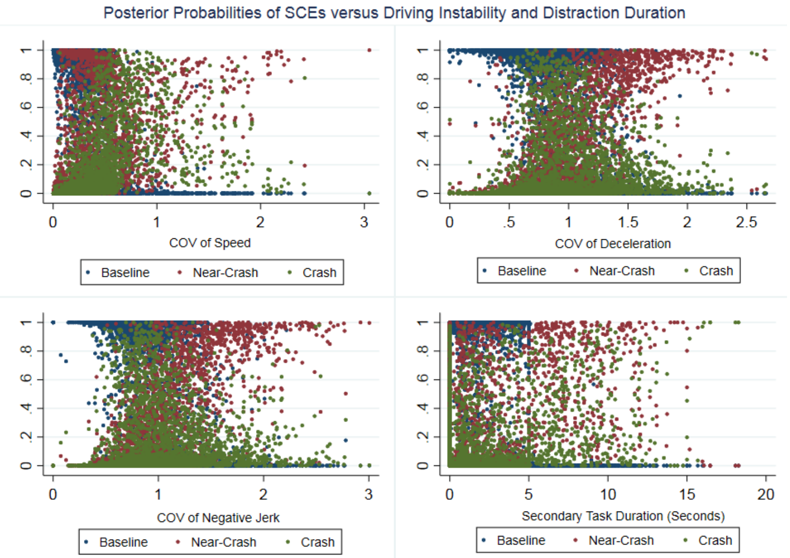
Modeling Results: Estimation of Ordered Probit Model

Key Explanatory Variables	Coeff	t-stats	Chance in % (Marginal Effects*100)		
Measures of Volatility			**Baseline**	**Near-Crash**	**Crash**
CV of Speed	0.956	9.97	-11.74	6.57	5.17
CV of Deceleration	1.107	13.25	-13.59	7.61	5.98
CV of Negative Jerk	0.214	2.76	-2.63	1.47	1.16
Secondary Task Duration (seconds)	0.032	4.18	-0.39	0.22	0.17
Mean Travel Speed (miles per hour)	-0.022	-4.96	0.26	-0.15	-0.12
Driving Errors (Base = no driving error)					
Recognition Error	1.681	22.49	-20.64	11.55	9.08
Decision Error	1.867	26.13	-22.93	12.84	10.09
Performance Error	2.544	11.91	-31.24	17.49	13.75
Violation	1.009	13.29	-12.39	6.94	5.45
Physical Condition-related Error	0.978	6.19	-12.01	6.72	5.29
Experience or Exposure Error	1.781	5.87	-21.86	12.24	9.62
Intersection Influence (Base = no intersection influence)					
Interchange Influence	0.785	7.76	-9.64	5.39	4.24
Stop Sign or Traffic Signal Influence	0.247	4.03	-3.03	1.70	1.33
Uncontrolled Intersection Influence	0.593	6.38	-7.28	4.07	3.20
Parking Lot or Driveway Entrance/Exit	0.755	8.80	-9.27	5.19	4.08
Other Intersection Influence	0.560	3.59	-6.87	3.85	3.02
Roadway Environment					
Urban area indicator (1/0)	0.319	3.08	-3.92	2.20	1.73
Moderate Residential (1/0)	-0.203	-3.32	2.50	-1.40	-1.10
Threshold parameters					
μ_1	2.801	27.92	---	---	---
μ_2	4.292	37.42	---	---	---
Summary Statistics					
N	6,464				
Pseudo R^2	0.493				
Loglikelihood (null)	-4013.632				
Loglikelihood (convergence)	-2036.74				
AIC	4113.479				
BIC	4248.959				

Machine Learning Results

Performance Measure	Ordered Probit	NB	KNN	GBT
Overall Accuracy (%)	85.75	89.75	88.70	91.23
Baseline				
Recall (%)	98.32	96.95	97.13	98.27
Precision (%)	92.07	95.65	94.39	96.21
F1 Score (%)	95.09	96.29	95.74	97.23
Near-Crash				
Recall (%)	38.93	68.70	61.83	71.50
Precision (%)	56.04	71.43	65.50	74.54
F1 Score (%)	45.95	70.04	63.61	72.99
Crash				
Recall (%)	34.44	47.78	44.44	48.33
Precision (%)	41.06	52.12	57.97	58.39
F1 Score (%)	37.46	49.86	50.31	52.89





Key Findings

- 5.17%, 5.98%, and 1.16% higher chance of a crash due to a unit increase in the COV of speed, COV of deceleration, and COV of a negative jerk, respectively.

- Performance error leads to the highest increase (13.75%) in crash risk followed by decision error (10.09%) and recognition error (9.08%)

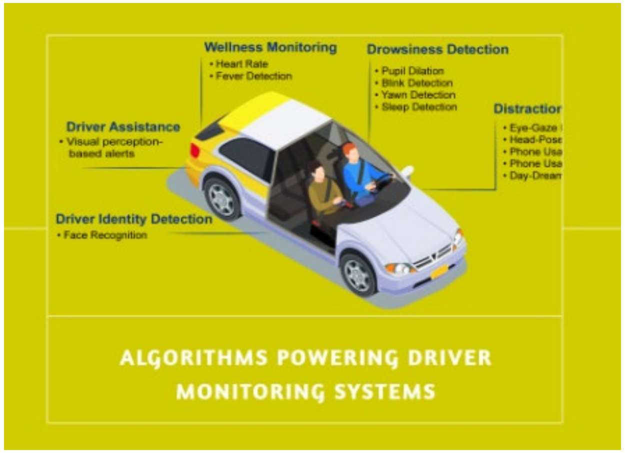
- Crash risk increases by 0.17% due to a unit increase in the duration of distraction while keeping other variables at their mean values.

- GBT classifier accurately predicts the event outcome (baseline, near-crash, and crash) in the test data with the highest prediction accuracy of 91.23%.

Future Research

The study findings advocate the need for:

- Training proactive ML-based algorithms which can warn drivers about the potential risk of SCEs in real-time based on driver behaviors and performance
- Developing ML algorithms, e.g., GBT classifier can be used to collect and process information from sensors in vehicles
  - Can start monitoring driving errors, violations, instability in driving, and duration of distraction to predict crash risk in real-time



Key Takeaways from the Project

Can driver distraction be identified in different driving scenarios?

- Driver distraction can be readily identified from instantaneous variations in driver biometrics, vehicle kinematics, and roadway surroundings.
- The volatility measures in driver biometrics, vehicle kinematics, and roadway surroundings associate with higher levels of distraction.
- The coefficients of variation in driver eye movements, distance of the vehicle from the lane centerline and the following vehicle were determined as key predictors of driver distraction.

Key Takeaways from the Project

How can driving events be classified as normal and distracted/impaired?

- Driving events can be classified as undistracted and distracted under various levels based on volatility measures
  - These include instantaneous variations in driver biometrics, vehicle kinematics, and roadway surroundings

- Coefficients of Variation in driver eye movements, vehicle speed, vehicle distance from the lane centerline, front and the following vehicle can classify driving events.

Key Takeaways from the Project

How is prolonged distracted driving associated with safety-critical events?

- Longer duration of distracted driving is significantly associated with instability in driving (volatility) which in turn leads to safety-critical events.
- Distraction duration is substantially longer in crashes than near-crashes and baselines.
- The probability of a crash increases exponentially when the distraction duration exceeds 8 seconds.

Key Takeaways from the Project

What are the mechanisms for driving errors and violations that lead to safety-critical events?

- Driving errors, classified into 5 types and violations are significantly associated with higher instability in driving speed (COV of speed).
- A unit increase in the COV of speed (measure of driving instability) increases the risk of crashes and near-crashes by 17.94% and 29.50%, respectively.
- All driving errors and violations not only contribute to SCEs directly but also indirectly through instability in driving speed.

THANK YOU