A Simulator Dataset to Support the Study of Impaired Driving

John Gideon*, Kimimasa Tamura, Emily Sumner, Laporsha Dees, Patricio Reyes Gomez, Bassamul Haq, Todd Rowell, Avinash Balachandran, Simon Stent, Guy Rosman

***Abstract*—Despite recent advances in automated driving technology, impaired driving continues to incur a high cost to society. In this paper, we present a driving dataset designed to support the study of two common forms of driver impairment: alcohol intoxication and cognitive distraction. Our dataset spans 23.7 hours of simulated urban driving, with 52 human subjects under normal and impaired conditions, and includes both vehicle data (ground truth perception, vehicle pose, controls) and driver-facing data (gaze, audio, surveys). It supports analysis of changes in driver behavior due to alcohol intoxication (0.10% blood alcohol content), two forms of cognitive distraction (audio n-back and sentence parsing tasks), and combinations thereof, as well as responses to a set of eight controlled road hazards, such as vehicle cut-ins. The dataset will be made available at https://toyotaresearchinstitute.github.io/IDD/.**

I. INTRODUCTION

Human drivers are prone to decision-making failures while driving, due to a range of behavioral impairments. Such failures can have significant consequences for human life. Two common sources of behavioral impairment for human drivers are alcohol intoxication and cognitive distraction.

Alcohol intoxication, even at moderate levels, is known to significantly increase the risk and severity of road traffic accidents. According to a recent report from the World Health Organization, about 20% of fatally injured drivers in high-income countries have blood alcohol concentration (BAC) levels above the legal limit, while in low- and middle-income countries that number rises to between 33% and 69% [1].

Cognitive distraction (CD) – inattention to the task of driving as mental processes are diverted to other activities – is far more prevalent in the driving population but poses relatively lower risk since it describes a transient mental state rather than lasting impairment. Distracted driving is typically divided into three types – visual (eyes off the road), manual (hands off the wheel), and cognitive (mind off the task) [2]. In the US, distracted driving contributes to around one-tenth of road traffic accidents [3]. While visual or manual distraction accounts for a lot of these incidents, cognitive distraction is the most difficult form of distraction to observe and measure [2], making mitigation challenging.

The contributions of our paper are as follows:

All authors are with the Toyota Research Institute, USA. Email: `firstname.lastname@tri.global`. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of TRI or any other Toyota entity.

* Corresponding author.

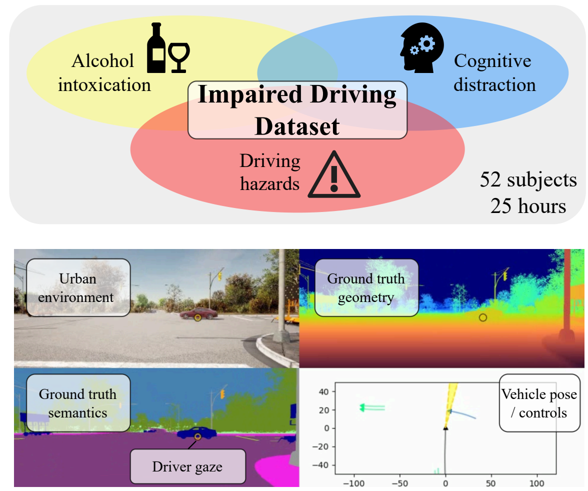

Fig. 1: **Impaired Driving Dataset overview.** We captured data from 52 human drivers over 25 hours of urban driving in a driving simulator, including under *alcohol-intoxicated* and *cognitively-distracted* driving conditions and with a range of realistic *driving hazards*. Our dataset supports various analyses including: the overlap between different types of impairment, how to distinguish one from the other, and their impact on behaviors such as visual attention and responses to road hazards.

- We design and release the first publicly available driving simulator experiment that combines cognitive distraction (CD), alcohol intoxication, and road hazards, as illustrated in Fig. 1.
- We run our experiment on 52 individuals, capturing a wide range of data including vehicle control during driving, eye gaze, ground truth scene state, ground truth impairment conditions, and various self-reported measures of driver state.
- We validate our data by analyzing the relationship between several well-defined behavioral driving features and the various impairment conditions.
- We establish a set of machine learning baselines using these features over four tasks – CD detection, intoxication detection, general impairment detection, and the ability of the model to differentiate types of impairment.

By releasing our dataset to the community, we hope to encourage further work to develop systems that can quickly and accurately diagnose impaired driving in various forms, in the pursuit of improved road safety. The dataset will be made available at https://toyotaresearchinstitute.github.io/IDD/.

TABLE I: **A summary of closely related datasets for studying simulated manual driving under impairment.** Our dataset is the first to explore the combination of alcohol intoxication, cognitive distraction and driving hazards, while providing coverage of driver gaze, controls and scene state. *One hazard scenario involving unintended acceleration. †Gaze derived from RGB video. ‡Cognitive workload (not distraction) is measured as a dependent variable.

	Variables of interest			*Sensor data*			*Dataset characteristics*		
Dataset	Alcohol	Cog. Dist.	Hazards	Gaze track	Controls	Scene ground truth	Subjects	Dur. (hr)	Scenarios
C42CN [4]	-	✓	✓*	✓	✓	-	68	80	Highway
CoCAtt [5]	-	✓	-	✓	✓	-	11	12	Countryside
Koch et al. [6]	✓	-	-	✓	✓	-	30	15	Varied
Keshtkaran et al. [7]	✓	-	-	-†	-	-	60	30	Urban
CL-Drive [8]	-	✓‡	-	✓	-	-	21	10	Varied
Ours	✓	✓	✓	✓	✓	✓	52	25	Urban

II. RELATED WORK

A. Alcohol intoxication and driving

The physiological and behavioral changes exhibited by drivers under the influence of alcohol have been well studied and documented. These changes include a decreased capacity to process visual information [9], an increased propensity to become distracted or not cope well with distractions [10], a drop in vigilance [11], and changes in visual behavior and eye movements linked to steering [12]. For a detailed review of these changes, please see [13].

While alcohol impairment in the driving context is typically measured using field sobriety tests administered during roadside stops, various attempts have been made to harness vehicle information to predict driver intoxication in real-time. Previous work has shown that it is possible to achieve performance comparable to the standardized field sobriety tests using eight minutes of driving observations [14], although differences between drivers and roadway situations had a large influence on algorithm performance. Using further information from a driver-facing camera (to extract eye, gaze and head movement features), it has been shown that blood alcohol concentration can be predicted in a laboratory setting to reasonable accuracy [6], [7]. In our study, we extend these prior efforts to more diverse driving conditions, incorporating realistic road hazards and cognitive distractions. Additionally, we make the raw data available to support future analyses.

B. Cognitive distraction and driving

Cognitive distraction has been extensively explored in the driving literature [15], [16], in terms of underlying psychological and physiological phenomena [17], [18] and its impact on behavior and risk [19]–[21]. It has been investigated for its interactions with other phenomena such as visual distraction [22], alcohol impairment [10], as well as its effect at different age groups [23].

Another thread of research involved possible approaches and interventions to address CD [24]–[27]. Several approaches have been used to detect both visual distraction and CD in various conditions and dataset realism levels [28]–[32]. However, these approaches were often explored within individual studies, making comparing them or reproducing them difficult, meriting a more comprehensive and unified dataset to serve as a benchmark within the research community

C. Behavioral driving datasets

Numerous on-road datasets exist for studying driver behaviors such as attention [33], gaze [34], in-cabin activities [35], or maneuver intent [36]. Studying the behaviors of drivers under impaired conditions presents a greater safety risk, hence the vast majority of datasets use driving simulators.

Several datasets that specifically address distraction have been released over the years [4], [5], [8]. We compare the most relevant datasets to ours in Table I. In general, such datasets are often limited to distraction alone, rather than including, e.g. interaction with alcohol impairment or curated interactions with specific road hazards. Moreover, many of them do not include vehicle controls or scene ground-truth information, which limits the features and patterns that can be observed and analyzed.

Our dataset is the first to mix the conditions of road risk and driver state. This makes it possible to study the joint and marginal effects of alcohol impairment and cognitive distraction during free driving and in response to realistic road hazards. We hope this dataset will alleviate some of the limitations of existing datasets towards reproducible and comprehensive research and evaluation of approaches for detecting cognitive distraction and intoxication, along with their interaction with both normal and hazardous driving conditions.

III. METHOD

A. Experimental procedure

To understand the effects of CD and alcohol intoxication, we collected two versions of the study: impaired driving with alcohol intoxication and CD (version 1, **v1**) and driving only with CD (version 2, **v2**). The overall experimental flow is visualized in Fig. 2.

v1 participants, who went through the intoxication procedure, were full-time Toyota Research Institute employees, while **v2** participants were externally recruited through User Interviews [38]. All participants signed a consent form prior to participation. The experimental procedure was reviewed and approved by the WCG IRB (Protocol #20241945). All personally identifiable information is stored on HIPAA-compliant machines to ensure participant privacy.

The study consisted of five primary components: 1) gaze calibration, 2) pre-driving visuomotor tasks, 3) driving practice, 4) CD task practice, and 5) driving with hazards. Participants were also given questionnaires at different points in the

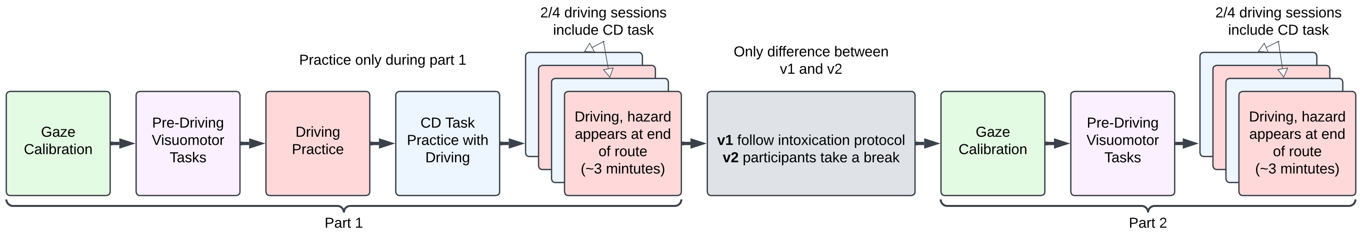

Fig. 2: **Experimental flow overview,** consisting of two main parts separated by a break. The two experiment versions (v1 and v2) only differ in that v1 participants follow the intoxication protocol in Sec. III-B and consume alcohol during the break.

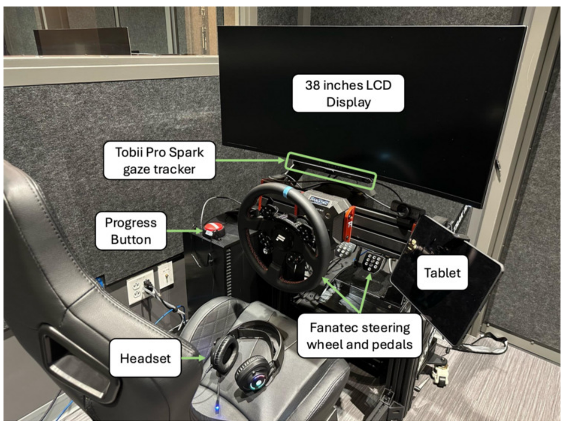

Fig. 3: **Driving simulator setup,** including steering wheel, pedals, tablet, and progress button to advance through the study. Driver-facing sensors included a Tobii Pro Spark eye tracker [37], headset microphone, and a webcam. Participants sat approximately 65 cm from the screen.

experiment. Each participant went through these components twice. After completing it the first time, **v1** participants followed the intoxication procedure while **v2** participants took a break. All study activity, with the exception of the intoxication procedure, took place in the driving simulator (Fig. 3).

Prior to driving, we calibrated participants' gaze tracking using the Tobii API [37]. Participants then spent 10 minutes performing some simple pre-driving visuomotor tasks as part of another experiment [39]. Next, participants began the driving portion of the study, which was performed in the CARLA [40] simulator's Town 15 map. Participants drove through a set route around the town for two minutes. Afterward, participants were introduced to each of the CD tasks: the 1-back [41] and sentence comprehension task [42]. Participants would practice each task twice, once while they were not driving and once while they were driving. Participants only practiced the CD tasks during the first part of the experiment.

Participants then drove a series of four scenarios of a set distance lasting approximately three minutes. Each of these ended with a hazard (depicted in Fig. 4) meant to induce a reaction in the driver. These hazards were designed to generate interesting behaviors and test different aspects of driving, such as reaction time and lateral scene awareness. We use scripted hazards to ensure reproducibility and comparability between all participants. The hazard logic was designed to allow flexibility to different driver speeds and to adjust the respective pedestrian or other vehicle speed accordingly.

Out of these four scenarios per block, two were carried out with no cognitive distraction and two were carried out with the added 1-back and sentence comprehension tasks (one each). In both cognitive distraction tasks, participants listened to audio through the headset and their spoken responses were recorded. In the 1-back task, a single-digit number was played every 2.5 seconds, and participants were asked to remember and respond with the previously presented number. In the statement task, participants listened to a short sentence of varying length and responded *subject*, *object*, and *Yes or No* about the plausibility of the sentence. This resulted in four routes driven per block, totaling eight routes driven and eight hazards across both parts per participant.

We employed four different orders in which participants experienced scenarios to ensure that each hazard had an even balance of baseline, CD tasks, and intoxication. After participants completed the first four scenarios, they began the intoxication portion (**v1**) or took a break (**v2**).

For each participant, the dataset includes several questionnaires from different time points in the experiment. We collected short-form PANAS [43] questionnaires before, and after each part, as well as NASA-TLX [44] and Karolinska Sleepiness Scale (KSS) [45] after each hazard.

B. Intoxication Procedure

v1 participants were instructed to not eat or drink anything for at least three hours before their study start time. Before beginning the experiment, they provided their weight and confirmed that they had a blood alcohol level (BAC) of 0 through a breathalyzer. The participant's weight and gender were used to determine how many grams of alcohol they needed to reach a BAC of 0.1. This threshold was selected, as prior work has shown an exponential increase in risk after a BAC of 0.1 [46]. Participant BAC levels were monitored throughout the study, including before and after the intoxicated portion.

Upon completion of the first part of the study, the proctor measured the correct amount of alcohol using a kitchen scale. Participants chose between tequila, vodka, or whiskey and were provided 1 oz of juice as a mixer or chaser. Participants had 10 minutes to consume all or as much of the alcohol as they could. The proctor assured participants that they were not required to drink all of the alcohol and could end the

(a) Pedestrian sudden crossing

(b) Pedestrian sudden crossing

(c) Child running out into road

(d) Car pop-out from parking spot

(e) Car door opening

(f) Car running stop sign merge

(g) Car running red light

(h) Car running red light

Fig. 4: **The dataset contains a wide variety of aligned hazardous events,** allowing for further investigation on the impact of impairment on driving risk.

study at any point. Afterwards, BAC levels were checked and recorded every 10 minutes. Once a participant had a BAC over 0.1 and their BAC was decreasing, participants went back to the simulator room to complete the second half of the study. If a participant had a BAC under 0.08, the wellness proctor would ask if the participant would be comfortable consuming an additional 20g of alcohol. If a participant said yes then the additional alcohol would be provided.

Once participants completed the impaired driving portion of the study, the proctor escorted participants to a recovery space, where they remained until reaching a BAC level appropriate for their return to work or home. The participants had an average peak BAC of 0.110±0.024 and average BAC of 0.085±0.017 immediately after the intoxicated driving portion.

IV. DATASET

A. Participants

In **v1** (n = 20), ages ranged from between 23-45 with the average age being 32.2 (SD 6.4). Thirteen participants identified as male, six identified as female, and one identified as non-binary. Ethnically, most identified as White (8) or Asian (7), leaving those who identified as Black/African American (1), Native Hawaiian or Other Pacific Islander (1), and mixed-raced (2) in the minority (one participant did not report their race/ethnicity and another did not report any of their demographics). There were initially 22 participants, 1 participant was excluded due to participating in a pilot with slightly different procedures. The other participant dropped out due to over intoxication. Despite targeting a BAC of 0.1, the calculation is only an estimate and each participant reacted differently to the amount ingested.

In **v2** (n = 32) ages ranged from 23 to 65, with a mean of 37.1 (SD 10.3). This portion included 18 participants who identified as male, 13 as female, and 1 as non-binary. A total of 9 participants identified ethnically as White, 9 as Black/African American, 7 as Asian, and 7 as Hispanic/latino. Of the initial 38 **v2** participants, 6 were excluded for dropping out of the study due to motion sickness.

B. Dataset Structure

We collected an overall dataset of 23.7 hours driving, for a total of 839 km, and a total of 395 hazard instances under various conditions of impaired/normal driving – see Table II for the full breakdown of driving data collected and Fig. 5 to see alignment within a single hazard.

In terms of scene information, forward-facing virtual RGB, depth, and semantic segmentation camera from CARLA's virtual sensors (sampled at 10 Hz and a 960×400 resolution) were recorded. CARLA ego state and controls were collected at 10 Hz, as well as and the other vehicle tracks. The other vehicle tracks were collected at 10 Hz from the CARLA state, with some impact (70/416 trips) due to an unknown simulator traffic spawning issue, which we annotated as unavailable. These state and scene logs were combined with gaze information — for gaze logging, Tobii Spark Pro [37] measurements were recorded at 60 Hz, and contained gaze vectors for each eye, gaze point on screen, and pupil diameter.

TABLE II: **Amount of data in each driver state.**

Driver State					
Alcohol	CD	Trips	Hours	Distance in km	Valid Hazards
No	No	168	9.9	342	160
No	Yes	168	9.7	336	158
Yes	No	40	2.1	81	40
Yes	Yes	40	2.1	81	37
Total		416	23.7	839	395

Fig. 5: **A montage of screenshots from a subset of participant experiences of the "pedestrian sudden crossing" hazard.** Hazards were triggered depending on each participant's speed leading up to the hazard location. Trigger logic was tuned to provide the most similar experience within normal speed bounds, allowing for direct comparisons of hazard responses across participants. We labeled 5% (22/416) of cases where the hazard did not deploy correctly, due to the participant driving off-route or excessively speeding, to remove them from any analyses.

TABLE III: **A breakdown by driver state of ego-vehicle collisions (crashes) incurred.** Middle: only counting collisions during one of the eight designed road hazards. Right: counting all collisions. Alcohol more than doubles the rate of collisions, while cognitive distraction only increases it marginally.

Driver State		*Designed Hazard*		*All Crashes*	
Alcohol	CD	Crash	Crash/Hazard	Crash	Crash/Trip
No	No	10	0.06	48	0.29
No	Yes	10	0.06	52	0.31
Yes	No	5	0.13	26	0.65
Yes	Yes	8	0.22	26	0.65
Total		33	0.08	152	0.37

C. Collisions

Collisions with other vehicles and pedestrians occurred throughout the collection of the study - both due to planned hazard interactions and emergent encounters mostly due to the dynamics of the lead vehicle. Table III shows the distribution of these crashes across the various driver states. While there was little effect seen for sober CD, intoxicated participants had a substantially higher collision rate of 0.13 versus 0.06 in designed hazards and 0.65 versus 0.29 throughout all driving.

D. Simulation Anomalies

We relied on randomly generated traffic patterns during the part of the experiment preceding the hazard both to allow for a wider variety of interactions and to make larger numbers of vehicles feasible to implement. However, this occasionally produced anomalous behaviors - such as traffic coming to a complete standstill. Because of this, once the data collection was complete, we visually inspected the entire dataset to generate temporal annotations for anomalous traffic events, in case it becomes necessary to remove them in future research.

Mistriggered hazards: Due to likely physics issues or participants driving too quickly, some of the hazards did not trigger as planned (as seen in Table II) and are annotated.

Despawning vehicles: In initial piloting it was found that participants could get into traffic deadlocks. To mitigate this, we despawn any vehicle that remains still for more than five seconds. This sometimes results in the disappearance of vehicles within view of the driver, especially near intersections. A list of these despawns and causes are included with the dataset.

Vehicle Displacement: An unknown simulation bug sometimes caused a stationary car to appear slightly off the road, affecting 1.5% of captured driving data. The start and end times of these occurrences are annotated. We remove all scenarios containing these anomalies from our analyses.

Participant off route: A total of eight participants drove substantially off the specified route during one of their scenarios. Of these, six returned to complete the route. The times of these off-route portions are annotated in the dataset.

Other vehicles driving through barriers: We used a set of traffic barriers to define the route to follow. However, since we wanted the other vehicles to be able to drive other routes, we allowed for them to drive through these barriers.

V. DATASET VALIDATION

We next demonstrate the ecological validity of our dataset for the study of alcohol intoxication and cognitively distracted driving by analyzing sets of well-known behavioral features established in prior work. We consider both control and gaze input modalities. When considering the gaze modality, we take the average of the aggregate features in the left and right eyes.

For each feature, we first calculate its value in all scenarios. For consistency, we only consider the 90 seconds of data that begins 120 seconds before the end of the scenario. This excludes the final 30 seconds to only focus on driving without hazards. We then consider three different driver state comparisons for analysis - (i) baseline (instances without either form of impairment) versus intoxication, (ii) baseline versus CD, and (iii) baseline versus both. For each comparison, we group the scenarios based on the conditions of the experiment and take the mean of the feature. We then perform a Wilcoxon signed-rank test to determine if there is a significant difference between the features in the two conditions. Because only a subset of participants were intoxicated during the experiment, only 20 participants are used for the (i) and (iii) comparisons, while all 52 are used for (ii). We report both the p-value of this test and the difference of the means in Table IV.

A. Speed and Acceleration

The disinhibition caused by alcohol has been found to correlate with higher speeds and accelerations when driving [47]. However, there is not yet a consensus in the field for the impact of CD on driving performance [48]. In order to calculate vehicle dynamics at a particular instant, we consider the 5-second window centered on the current time. We then fit a cubic spline weighted with a Hann window and calculate the longitudinal speed and acceleration. Our results show mean speed in meters per second (m/s) and mean accelerations in meters per second squared (m/s^2).

Our analysis did not find a significant impact of impairment on speed. Both types of acceleration were found to be significantly higher in intoxication, as in [47], regardless of the presence of CD. CD had a mixed impact on acceleration – significantly lowering forward acceleration when compared with baseline. This interfered with the effect of intoxication on acceleration and resulted in the increase no longer being significant when both states occurred simultaneously.

B. Steering Reversals

The steering reversal rate (SRR) aims to capture the amount of steering corrections needed by a driver to follow a route. Kountouriotis et al. noted an increase in SRR when drivers experienced different forms of CD [48], while Li et al. found a similar increase during intoxication [49]. We follow the work of Markkula et al. to calculate SRR [50] and select thresholds of 0.5 and 2.5 degrees, as in [48]. We report the counts of SRR per minute as the feature.

Intoxication resulted in a significant increase in both types of corrections versus baseline, as seen in [49]. However, CD did not have a significant difference and resulted in

Feature	Baseline vs Intox.			Baseline vs CD			Baseline vs Both		
	Diff.	Effect	p-value	Diff.	Effect	p-value	Diff.	Effect	p-value
Longitudinal Speed (m/s)	0.267	0.092	0.6813	-0.101	0.049	0.7225	1.107	0.409	0.0674
Longitudinal Acceleration (m/s^2)	0.114	0.559	0.0124	-0.066	0.418	0.0026	0.056	0.209	0.3507
Braking Acceleration (m/s^2)	0.128	0.526	0.0187	-0.033	0.164	0.2365	0.127	0.442	0.0479
Steering Reversals at 0.5 Degrees (#/min)	3.667	0.693	0.0019	0.549	0.147	0.2883	3.700	0.635	0.0045
Steering Reversals at 2.5 Degrees (#/min)	1.317	0.545	0.0149	-0.141	0.191	0.1690	0.900	0.397	0.0760
Gaze Pitch Standard Dev. (Degrees)	-0.275	0.459	0.0400	-0.671	0.741	0.0000	-0.666	0.760	0.0007
Gaze Yaw Standard Dev. (Degrees)	-0.644	0.342	0.1259	-0.494	0.422	0.0024	-0.706	0.442	0.0479
Saccades (#/min)	-58.150	0.877	0.0001	-26.849	0.571	0.0000	-72.633	0.860	0.0001
Fixations (#/min)	-45.067	0.877	0.0001	-25.031	0.687	0.0000	-60.017	0.868	0.0001
Pupil Diameter (mm)	0.003	0.025	0.9108	0.080	0.615	0.0000	0.136	0.442	0.0479

TABLE IV: The results of our dataset validation. We consider four different pairwise comparisons and report their distribution mean differences and the effect size and p-value calculated using a Wilcoxon signed-rank test. We gray out any result with a p-value greater than 0.05 as not significant.

mixed effects when considering with intoxication jointly. The minor steering reversals at 0.5 degrees were found to occur at a slightly higher rate when jointly modeled than when intoxicated alone, potentially indicating some effect, as in [48].

C. Gaze Pitch and Yaw Standard Deviation

Prior work has noted that gaze becomes more concentrated on the road when cognitively loaded [48] or intoxicated [12]. One method to measure this is to take the standard deviation of the yaw of the gaze vector, as in [48], which should decrease as the gaze becomes more concentrated. We report both gaze pitch and yaw standard deviation in degrees.

We noted a significant decrease in gaze pitch standard deviation (increased concentration) in both intoxicated and CD experiments and yaw in CD experiments, as noted in [48] and [12]. A similar concentration was seen for both pitch and yaw when jointly modeled.

D. Fixation and Saccade Counts

Fixations are the periods when eye position remains relatively stable, allowing for detailed visual processing of an object or location, while saccades are the quick movements in between. Prior work has found that the occurrence of each of these decreases with both intoxication [51] and CD [52]. We use PyGaze [53] to extract the instances of fixations and saccades and obtain the counts per minute as a feature.

Both fixations and saccades significantly decreased in both driver states, as noted in prior works [51], [52]. The effect further increases when jointly modeled.

E. Pupil Diameter

Prior work has found that pupil size increases while engaged in a cognitive task [52], while it decreases while intoxicated [54]. We calculate the mean pupil diameter across each scenario and report the amount in millimeters (mm).

Our results show a significant increase in pupil size when cognitively distracted (as in [52]) and in mixed data, with no effect seen for intoxication alone. Further work is needed to investigate the multifaceted nature of pupil response time signals [55], [56], which can be explored in our dataset. This could include variability due to screen brightness and participant fatigue, which may need to be explicitly modeled for a more thorough future analysis.

VI. MACHINE LEARNING BASELINES

A. Machine Learning Methods

In this section, we explore the usefulness of the above set of 10 features at detecting CD and intoxication. We also consider instead framing the task as first detecting any impairment and then differentiating the particular type of impairment. For this last task, any data that is both CD and intoxicated is considered intoxicated for classification purposes.

We extract 30 second segments from the dataset without overlap. We begin extraction once 30 seconds has elapsed and end extraction 30 seconds before the end of the recording to remove the impact of edge effects and hazards. We then calculate the above feature for input to a variety of models. We employ a fixed set of five folds, divided by subject, which will be released with the dataset. Experiments are conducted in a round robin fashion, resulting in a set of five unweighted average recalls (UARs), which are then averaged.

We consider five main models. The first four are random forest, gradient boosting, a support vector machine, and logistic regression and are implemented using the default scikit-learn [57] parameters and balanced class weights. We also implement a shallow multilayer perceptron (MLP) in PyTorch [58] with three hidden layers of sizes 256, 128, and 64. Each hidden layer consists of the linear layer, Leaky ReLU activation, batch normalization, and dropout (p=0.35). We train using the AdamW optimizer and a learning rate of 1e-5. We consider both the cases of training four separate single task classifiers and a single multi-task classifier.

B. Machine Learning Results

Our results across all models can be seen in Table V. Overall, logistic regression and the multi-task MLP provide the best performance across tasks. The tasks that involve distinguishing intoxication from either baseline or CD have a greater range of model performances, implying that it may be a more difficult task than simply detecting general impairment. Using a multi-task MLP provides a slight improvement over all of the single-task MLPs, especially when differentiating

Model	CD	Intox.	Impaired	Diff.
Random Forest	0.64±0.02	0.62±0.07	0.67±0.03	0.63±0.07
Grad. Boosting	0.63±0.02	0.66±0.06	0.67±0.05	0.65±0.06
SVM	0.65±0.01	0.59±0.04	0.67±0.02	0.66±0.05
Logistic Reg.	**0.67±0.01**	**0.69±0.03**	0.68±0.03	**0.69±0.03**
Single-task MLP	0.64±0.01	0.67±0.05	0.68±0.04	0.63±0.08
Multi-task MLP	0.65±0.02	**0.69±0.05**	**0.69±0.05**	0.67±0.04

TABLE V: **The results of our machine learning baselines.** We explore six different models and determine the UAR on four tasks – CD, intoxication, general impairment detection, and differentiating (diff.) between types of impairment, given only impaired data.

		Predicted		
		Baseline	CD	Intox.
Actual	Baseline	70%	17%	13%
	CD	36%	43%	21%
	Intox.	25%	22%	53%

TABLE VI: **Confusion matrix for the multi-task MLP.** This model produces relatively low false intoxication rates of 13% for baseline and 21% for CD.

between impairment types. This provides evidence that similar mixed datasets and multi-task methods may be needed to be able to appropriately detect and react to drivers' mental state.

We next construct a 3-class confusion matrix (seen in Table VI) using the impaired and differentiating tasks from the multi-task MLP. We use this model and set of tasks, as it should provide a low false positive rate, due to its strong ability to separate impaired and baseline data. When considering actual baseline data, only 17% is mischaracterized as CD, while 13% is confused for intoxication. For tested CD data, 21% is predicted as intoxication.

VII. LIMITATIONS

While the dataset is comparable in participant count and duration to similar driver impairment-focused datasets [4]–[8], it is relatively small when compared with many (particularly autonomous) driving datasets. Regardless, we aim to continue to increase the size and diversity of this dataset in future work. One area for improvement will be to incorporate more realistic cognitive tasks. While 1-back and sentence compression tasks are typical CD baselines, they do not necessarily reflect typical in-vehicle behavior. Simulated driving datasets are also ultimately limited by the realism of the scene and the naturalness of the behavior. But due to safety concerns, it is difficult to collect a controlled experiment with in-vehicle intoxication, especially one containing risky scenarios [6], [7].

VIII. CONCLUSIONS

We have created the first combined driving dataset containing annotated examples of multiple forms of driver impairment, as well as their reactions to hazards in each state. We showed that our dataset has reasonable ecological validity by examining how various gaze and control-based features change under different types of impairment. We found that gaze features are generally useful for detecting both types of driver impairment, while control features are more effective for intoxication. These features tend to correlate to what has been found in prior work, validating that the dataset contains expected phenomena of intoxication and CD. We establish a set of machine learning baselines over a set of tasks aimed at detecting and differentiating different types of impairment. We find that differentiation is a more difficult task and that multi-task models may provide some benefit. Furthermore, our collision statistics (Table III) indicate that a better understanding of driver impairment could result in improved design of interventions, which could be directly tested using this dataset.

Our dataset supports a variety of research areas, including the prediction of (i) online / multi-task impairment, (ii) driver hazard response, (iii) speed profile, and (iv) gaze. In particular, each area would be of interest when conditioned on known impairment and driver gaze. We view our dataset as a complement to real-world data that allows for deeper exploration of the intersection between impairment and responses to hazardous events in a way never before publicly available.

REFERENCES

[1] Safety and Mobility, Social Determinants of Health teams, "Global status report on road safety," World Health Org., Tech. Rep., 2023.

[2] B. Hamilton and J. Grabowski, "Cognitive distraction: Something to think about," *AAA Foundation for Traffic Safety*, 2013.

[3] National Center for Statistics and Analysis, "Research note: Distracted driving in 2022," *National Highway Traffic Safety Administration*, vol. DOT HS 813 559, 2024.

[4] S. Taamneh, P. Tsiamyrtzis, M. Dcosta, P. Buddharaju, A. Khatri, M. Manser, T. Ferris, R. Wunderlich, and I. Pavlidis, "A multimodal dataset for various forms of distracted driving," *Sci. Data*, vol. 4, no. 1, p. 170110, Aug. 2017.

[5] Y. Shen, N. Wijayaratne, P. Sriram, A. Hasan, P. Du, and K. Driggs-Campbell, "CoCAtt: A cognitive-conditioned driver attention dataset," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 2022, pp. 32–39.

[6] K. Koch, M. Maritsch, E. Van Weenen, S. Feuerriegel, M. Pfäffli, E. Fleisch, W. Weinmann, and F. Wortmann, "Leveraging driver vehicle and environment interaction: Machine learning using driver monitoring cameras to detect drunk driving," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, Hamburg Germany, 2023, pp. 1–32.

[7] E. Keshtkaran, B. von Berg, G. Regan, D. Suter, and S. Z. Gilani, "Estimating Blood Alcohol Level Through Facial Features for Driver Impairment Assessment," *WACV*, 2024.

[8] P. Angkan, B. Behinaein, Z. Mahmud, A. Bhatti, D. Rodenburg, P. Hungler, and A. Etemad, "Multimodal brain–computer interface for in-vehicle driver cognitive load measurement: Dataset and baselines," *IEEE Trans. on Intell. Transp. Syst.*, vol. 25, no. 6, pp. 5949–5964, 2024.

[9] K. Ziedman, H. Moskowitz, and R. A. Niemann, "The effects of alcohol on the driver's visual information processing," DOT, Tech. Rep., Sept. 1980.

[10] C. Ahlström, R. Zemblys, S. Finér, and K. Kircher, "Alcohol impairs driver attention and prevents compensatory strategies," *Accident Analysis & Prevention*, vol. 184, p. 107010, 2023.

[11] H. S. Koelega, "Alcohol and vigilance performance: a review," *Psychopharmacology*, vol. 118, no. 3, pp. 233–249, Apr. 1995. [Online]. Available: https://doi.org/10.1007/BF02245951

[12] E. Tivesten, V. Broo, and M. L. Aust, "The influence of alcohol and automation on drivers' visual behavior during test track driving," *Transportation research part F: traffic psychology and behaviour*, vol. 95, pp. 215–227, 2023.

[13] M. Dong, Y. Y. Lee, J. S. Cha, and G. Huang, "Drinking and driving: A systematic review of the impacts of alcohol consumption on manual and automated driving performance," *Journal of Safety Research*, vol. 89, pp. 1–12, June 2024.

[14] J. D. Lee, D. Fiorentino, M. L. Reyes, T. L. Brown, O. Ahmad, J. Fell, N. Ward, and R. Dufour, "Assessing the feasibility of vehicle-based sensors to detect alcohol impairment," *Washington, DC: National Highway Traffic Safety Administration*, vol. 1, no. 2, p. 7, 2010.

[15] M. A. Regan, C. Hallett, and C. P. Gordon, "Driver distraction and driver inattention: definition, relationship and taxonomy," *Accid. Anal. Prev.*, vol. 43, no. 5, pp. 1771–1781, Sept. 2011.

[16] L. Yekhshatyan and J. D. Lee, "Changes in the correlation between eye and steering movements indicate driver distraction," *IEEE Trans. on Intell. Transp. Syst.*, vol. 14, pp. 136–145, 2013.

[17] K. Hurts, L. S. Angell, and M. A. Perez, "The distracted driver: Mechanisms, models, and measurement," *Reviews of human factors and ergonomics*, vol. 7, no. 1, pp. 3–57, 2011.

[18] J. D. Lee, "Dynamics of driver distraction: The process of engaging and disengaging," *Annals of advances in automotive medicine*, vol. 58, p. 24, 2014.

[19] E. D. Sussman, H. Bishop, B. Madnick, R. Walter, *et al.*, "Driver inattention and highway safety," in *Annual Meeting of the Transportation Research Board*. United States. National Highway Traffic Safety Administration, 1985.

[20] D. Ascone, T. Tonja Lindsey, C. Varghese, *et al.*, "An examination of driver distraction as recorded in nhtsa databases," United States. National Highway Traffic Safety Administration, Tech. Rep., 2009.

[21] R. Talbot, H. Fagerlind, and A. Morris, "Exploring inattention and distraction in the safetynet accident causation database," *Accident Analysis & Prevention*, vol. 60, pp. 445–455, 2013.

[22] Y. Liang and J. D. Lee, "Combining cognitive and visual distraction: Less than the sum of its parts," *Accident Analysis & Prevention*, vol. 42, no. 3, pp. 881–890, 2010.

[23] T. Tanaka, K. Fujikake, Y. Yoshihara, N. Karatas, K. Shimazaki, H. Kanamori, and H. Aoki, "Analysis of distraction and driving behavior improvement using a driving support agent for elderly and non-elderly drivers on public roads," in *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2020, pp. 1029–1034.

[24] J. D. Lee, J. D. Hoffman, and E. Hayes, "Collision warning design to mitigate driver distraction," in *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, 2004, pp. 65–72.

[25] M. N. Lees and J. D. Lee, "The influence of distraction and driving context on driver response to imperfect collision warning systems," *Ergonomics*, vol. 50, no. 8, pp. 1264–1286, 2007.

[26] C. Ahlstrom, K. Kircher, and A. Kircher, "A gaze-based driver distraction warning system and its effect on visual behavior," *IEEE Trans. on Intell. Transp. Syst.*, vol. 14, no. 2, pp. 965–973, 2013.

[27] T. Tanaka, K. Fujikake, Y. Yoshihara, N. Karatas, H. Aoki, H. Kanamori, *et al.*, "Preliminary study for feasibility of driver agent in actual car environment—driver agent for encouraging safe driving behavior (3)," *Journal of Transportation Technologies*, vol. 10, no. 02, p. 128, 2020.

[28] M. Kutila, M. Jokela, T. Mäkinen, J. Viitanen, G. Markkula, and T. Victor, "Driver cognitive distraction detection: Feature estimation and implementation," pp. 1027 – 1040, 2007.

[29] S. Yang, J. Kuo, and M. G. Lenné, "Analysis of gaze behavior to measure cognitive distraction in real-world driving," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 62, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2018, pp. 1944–1948.

[30] Q. Sun, C. Wang, Y. Guo, W. Yuan, and R. Fu, "Research on a cognitive distraction recognition model for intelligent driving systems based on real vehicle experiments," *Sensors*, vol. 20, no. 16, p. 4426, 2020.

[31] C. Ahlström, G. Georgoulas, and K. Kircher, "Towards a context-dependent multi-buffer driver distraction detection algorithm," *IEEE Trans. on Intell. Transp. Syst.*, pp. 1–13, 2021.

[32] A. Misra, S. Samuel, S. Cao, and K. Shariatmadari, "Detection of driver cognitive distraction using machine learning methods," *IEEE Access*, vol. 11, pp. 18 000–18 012, 2023.

[33] J. D. Ortega, N. Kose, P. Cañas, M.-A. Chao, A. Unnervik, M. Nieto, O. Otaegui, and L. Salgado, "Dmd: A large-scale multi-modal driver monitoring dataset for attention and alertness analysis," in *ECCV Workshops*. Springer, 2020, pp. 387–405.

[34] I. Kasahara, S. Stent, and H. S. Park, "Look both ways: Self-supervising driver gaze estimation and road scene saliency," in *ECCV*, 2022.

[35] D. Yang, S. Huang, Z. Xu, Z. Li, S. Wang, M. Li, Y. Wang, Y. Liu, K. Yang, Z. Chen, Y. Wang, J. Liu, P. Zhang, P. Zhai, and L. Zhang, "Aide: A vision-driven multi-view, multi-modal, multi-tasking dataset for assistive driving perception," in *ICCV*, 2023, pp. 20 459–20 470.

[36] A. Jain, H. S. Koppula, B. Raghavan, S. Soh, and A. Saxena, "Car that knows before you do: Anticipating maneuvers via learning temporal driving models," in *ICCV*, 2015, pp. 3182–3190.

[37] *Tobii Spark Pro User Guide*, 1st ed., Tobii, 2024, available at https://example.com/manual.pdf.

[38] "User interviews," https://www.userinterviews.com/.

[39] S. Stent, J. Gideon, K. Tamura, A. Balachandran, and G. Rosman, "Beyond breathalyzers: Towards pre-driving sobriety testing with a driver monitoring camera," in *Proceedings of the 36th IEEE Intelligent Vehicles Symoposium (IV)*, 2025.

[40] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *CoRL*, 2017, pp. 1–16.

[41] D. He, B. Donmez, C. C. Liu, and K. N. Plataniotis, "High cognitive load assessment in drivers through wireless electroencephalography and the validation of a modified n-back task," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 4, pp. 362–371, 2019.

[42] M. R. Baumann and J. F. Krems, "A comprehension based cognitive model of situation awareness," in *Digital Human Modeling: Second International Conference, ICDHM 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19-24, 2009. Proceedings 2*. Springer, 2009, pp. 192–201.

[43] E. R. Thompson, "Development and validation of an internationally reliable short-form of the positive and negative affect schedule (panas)," *Journal of cross-cultural psychology*, vol. 38, no. 2, pp. 227–242, 2007.

[44] A. Cao, K. K. Chintamani, A. K. Pandya, and R. D. Ellis, "Nasa tlx: Software for assessing subjective mental workload," *Behavior research methods*, vol. 41, pp. 113–117, 2009.

[45] T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *International journal of neuroscience*, vol. 52, no. 1-2, pp. 29–37, 1990.

[46] R. D. Blomberg, R. C. Peck, H. Moskowitz, M. Burns, and D. Fiorentino, "The long beach/fort lauderdale relative risk study," *Journal of safety research*, vol. 40, no. 4, pp. 285–292, 2009.

[47] M. T. Fillmore, J. S. Blackburn, and E. L. Harrison, "Acute disinhibiting effects of alcohol as a factor in risky driving behavior," *Drug and alcohol dependence*, vol. 95, no. 1-2, pp. 97–106, 2008.

[48] G. K. Kountouriotis, P. Spyridakos, O. M. Carsten, and N. Merat, "Identifying cognitive distraction using steering wheel reversal rates," *Accident Analysis & Prevention*, vol. 96, pp. 39–45, 2016.

[49] Z. Li, X. Li, X. Zhao, and Q. Zhang, "Effects of different alcohol dosages on steering behavior in curve driving," *Human Factors*, vol. 61, no. 1, pp. 139–151, 2019.

[50] G. Markkula and J. Engström, "A steering wheel reversal rate metric for assessing effects of visual and cognitive secondary task load," in *Proceedings of the 13th ITS World Congress*. Leeds, 2006.

[51] J. B. S. Silva, E. D. Cristino, N. L. d. Almeida, P. C. B. d. Medeiros, and N. A. d. Santos, "Effects of acute alcohol ingestion on eye movements and cognition: a double-blind, placebo-controlled study," *PLoS One*, vol. 12, no. 10, p. e0186061, 2017.

[52] K. Walter and P. Bex, "Cognitive load influences oculomotor behavior in natural scenes," *Scientific Reports*, vol. 11, no. 1, p. 12405, 2021.

[53] E. S. Dalmaijer, S. Mathôt, and S. Van der Stigchel, "Pygaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments," *Behavior research methods*, vol. 46, pp. 913–921, 2014.

[54] E. L. Jolkovsky, F. E. Fernandez-Penny, M. Alexis, L. N. Benson, B. H. Wang, and B. S. Abella, "Impact of acute intoxication on quantitative pupillometry assessment in the emergency department," *Journal of the American College of Emergency Physicians Open*, vol. 3, no. 5, p. e12825, 2022.

[55] S. Mathôt, "Pupillometry: Psychology, physiology, and function," *Journal of cognition*, vol. 1, no. 1, 2018.

[56] B. Zandi and T. Q. Khanh, "Deep learning-based pupil model predicts time and spectral dependent light responses," *Scientific reports*, vol. 11, no. 1, p. 841, 2021.

[57] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[58] A. Paszke, "Pytorch: An imperative style, high-performance deep learning library," *arXiv preprint arXiv:1912.01703*, 2019.