Improving Methods to Measure Attentiveness through Driver Monitoring

July 2022 | Final Report

SAFE-D
SAFETY THROUGH DISRUPTION

Texas A&M Transportation Institute

San Diego State University
Leadership Starts Here

VT Transportation Institute
Virginia Tech.

Disclaimer

TECHNICAL REPORT DOCUMENTATION PAGE

<table>
<tr><td>1. Report No.
05-091</td><td>2. Government Accession No.</td><td colspan="2">3. Recipient's Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
Improving Methods to Measure Attentiveness through Driver Monitoring</td><td colspan="2">5. Report Date
July 2022</td></tr>
<tr><td colspan="2">6. Performing Organization Code:</td></tr>
<tr><td colspan="2">7. Author(s)
Marty Miller
Eileen Herbers
Jacob Walters
Luke Neurauter</td><td colspan="2">8. Performing Organization Report No.
Report 05-091</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address:
Safe-D National UTC
Virginia Polytechnic Institute and State University
Virginia Tech Transportation Institute
3500 Transportation Research Plaza
Blacksburg, Virginia 24061 USA</td><td colspan="2">10. Work Unit No.</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
69A3551747115/05-091</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
Office of the Secretary of Transportation (OST)
U.S. Department of Transportation (US DOT)
Office of the Assistant Secretary for Research and Technology
University Transportation Centers Program
Department of Transportation
Washington, DC 20590 United States</td><td colspan="2">13. Type of Report and Period
Final Research Report</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code</td></tr>
<tr><td colspan="4">15. Supplementary Notes
This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.</td></tr>
<tr><td colspan="4">16. Abstract
Driver inattention poses a significant problem on today's roadways, increasing risk for all road users. This report details our efforts in developing algorithms to detect driver inattention. A benchmark dataset was developed based on video review of driving events. Buffer-based algorithms were developed and compared using this benchmark dataset. The benchmark events were also used as a training dataset for machine learning models. Driver glance locations were important for determining driver attentiveness. In addition, vehicle speed was important for understanding the driving context, which was found to have a large impact on driver behavior.</td></tr>
<tr><td colspan="2">17. Key Words
Algorithms, distraction, inattention, attentiveness, machine learning</td><td colspan="2">18. Distribution Statement
No restrictions. This document is available to the public through the Safe-D National UTC website, as well as the following repositories: VTechWorks, The National Transportation Library, The Transportation Library, Volpe National Transportation Systems Center, Federal Highway Administration Research Library, and the National Technical Reports Library.</td></tr>
<tr><td>19. Security Classif. (of this report) Unclassified</td><td>20. Security Classif. (of this page) Unclassified</td><td>21. No. of Pages
21</td><td>22. Price
$0</td></tr>
</table>

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Abstract

Driver inattention poses a significant problem on today's roadways, increasing risk for all road users. This report details our efforts in developing algorithms to detect driver inattention. A benchmark dataset was developed based on video review of driving events. Buffer-based algorithms were developed and compared using this benchmark dataset. The benchmark events were also used as a training dataset for machine learning models. Driver glance locations were important for determining driver attentiveness. In addition, vehicle speed was important for understanding the driving context, which was found to have a large impact on driver behavior.

Acknowledgements

The research team would like to thank General Motors for allowing us to use the database that enabled our research. We would also like to extend our thanks to Dan Glaser at GM for serving as our Industry Champion and providing guidance throughout the project. Thanks to Kevin Kefauver for his review of this report. A special thanks goes to Eric Glenn, Safe-D UTC coordinator, for his assistance throughout the project in managing our Safe-D deliverables and facilitating Safe-D seminars which were instrumental in our research.

This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.

Table of Contents

List of Figures

List of Tables

Introduction

In 2020, “as many as 42,060 people are estimated to have died in [U.S.] motor vehicle crashes” with a further 4.8 million serious injuries occurring. This resulted in approximately $474 billion in economic cost to society (NSC, 2021). Based on the Second Strategic Highway Research Program (SHRP 2) Naturalistic Driving Study, distracted driving alone was observed during 63.8% of crashes (Dingus et al., 2016). This shows the critical importance of limiting distracted driving and serves as the impetus for the project detailed in this report: to design better inattention detection algorithms using driver eye tracking and vehicle metrics.

In fact, Level 2 (L2) advanced driver assistance systems (ADAS) (automated cruise control [ACC] + lane keep assist [LKA]) have a tendency to exacerbate drivers’ eyes-off-road time (EORT), a common stand-in for inattention or distraction, according to a AAA report (Dunn, Dingus, and Soccolich, 2019). It is perhaps for this reason that when General Motors (GM) introduced Super Cruise, the first hands-off L2 system available, it made use of a driver monitoring system (DMS) to ensure that drivers were paying attention to the forward roadway. L2 ADAS have the potential to increase driver safety by acting as a safety net for human drivers. But this safety benefit is only fully realized if the driver remains attentive, and DMSs have an important role to play in ensuring that is the case.

In an attempt to mitigate the effects of distraction, the United Nations has initiated vehicle requirements for DMSs, starting in 2023. This will require all new vehicles manufactured for use in the European market to feature driver-sensing technologies starting that year. While this move has great potential for alleviating the impacts of distracted driving, it is not a given that all DMSs are adequate for accurately detecting inattention. If these systems are not accurate, then drivers may begin to ignore any potential warnings of distracted driving issued. For this reason, it is important that algorithms for detecting inattention be as accurate as possible to increase drivers’ trust in the system.

Through a previous privately funded project, the research team had access to a naturalistic dataset housing DMS data along with other vehicle parameters, such as brake, throttle, and steering wheel position. The goal for this project was to leverage this database to investigate algorithms to distinguish between attentive and inattentive driver states.

Background

Distracted driving is difficult to confidently detect due to wide variation in driver behavior across driving environments and contexts. The most straightforward determination of distracted driving is when a driver’s eyes leave the road. A study using data from the 100-Car Naturalistic Driving Study found that glances totaling more than 2 seconds off-road increased the crash/near-crash risk by at least two times (Klauer et al., 2006). However, there are occasions where drivers may be looking at the forward roadway but are cognitively distracted (i.e., drowsy or daydreaming)

(Masala et. al, 2014). Conversely, an attentive driver making a sharp turn or driving in high pedestrian traffic may be scanning other areas around the vehicle besides the forward roadway.

Although DMSs are becoming more prevalent—in part to identify distracted, impaired, or otherwise disengaged drivers—the optimal approach for identifying these behaviors is still up for debate. This report will be focused on detecting driver inattention. The most direct way to determine driver inattention is by analyzing the driver's glance behavior using eye-tracking data. Previous algorithms use combinations of different glance patterns like glance location, glance frequency, glance duration, and eyelid movement (i.e., blinking patterns) (Kircher and Ahlstrom, 2018). For example, the SAVE-IT project developed an algorithm that considers the proportion of off-road glances during the last 3 seconds and the current off-road glance duration (Smith et. al, 2008). A study researching the eye-movement measures to in-vehicle tasks determined the degree of distraction by analyzing the percentage of time the driver's gaze is fixated on the road center during the last minute (Victor et al., 2005). The AttenD algorithm uses a time buffer that is depleted when the driver looks away from the road and reset when the driver looks back at the road for a set amount of time (Kircher and Ahlstrom, 2009). An experiment conducted by Smart Eye found that gaze fixation away from the forward roadway decreased reaction time more than adding an additional cognitive task, so their algorithm determines distraction exclusively from gaze fixation away from the forward road (Smart Eye, 2021).

Additionally, including other vehicle parameters and driver behavior measures can create a more robust algorithm when determining a driver's attention state. Driver facial affects, lane keeping, steering movements, time headway, and pedal movement were all used in a simulation study to determine distraction (McCall and Trivedi, 2004). For example, the frequency of steering correction decreases for inattentive drivers, while the magnitude of corrective action increases. Additionally, knowing a driver's intended actions can help predict their glance pattern. This is especially important in high-traffic areas and during turning maneuvers since the driver's glance may not be directed to only the forward roadway. Designing a driver model to predict driver behavior is one way to do so (Lie and Salvucci, 2001). By identifying an attentive driver model, any deviation from this model could indicate that a driver is inattentive (Wang et al., 2014).

Most previous algorithms focused on either using an intuitive approach to model the behavior of an attentive driver or using machine learning techniques to determine a driver's attention state. Machine learning techniques can be used to objectively decide how different glance patterns and metrics determine distraction. Long short-term memory (LSTM) recurrent neural networks use memory cells to store and access information over long time periods (driving events) to determine driver distraction using subject-specific data (Wollmer et al., 2011). Tango et al. (2013) found that a support vector machine (SVM) outperformed other machine learning techniques in determining driver distraction. SVMs are supervised learning models with algorithms that use regression analysis to classify data.

Although considerable attention has been paid to the topic of detecting driver inattention, a standardized way to detect this behavior using currently available data and methodologies has not emerged. Previous studies have indicated that including a combination of vehicle parameters and glance behavior can be the best predictor of driver attention. Thus, this research report seeks to investigate which methods and combinations of data are best used in determining driver inattention.

Methods

Naturalistic Database

This project had access to two naturalistic driving datasets, both captured during privately funded field operational tests, to use for designing and refining attention algorithms. These databases were collected during proprietary studies funded by GM and feature kinematic variables, such as vehicle speed/acceleration along with driver inputs (steering, brake, and throttle), and driver monitoring data. The DMS data consist of coded locations of where the driver was looking (Forward, Center Stack, Side Mirror, etc., or On Road/Off Road depending on the database) for every timestep. Although different DMSs were used for each of the datasets, both consisted of infrared cameras mounted on the steering column with infrared illuminators to ensure that drivers' faces were visible to the camera. Each DMS provided predictions of the driver's glance location throughout all drives at a rate of 10 Hz or greater.

All data was recorded asynchronously at varying time intervals, but generally in the 10 Hz or greater range. As part of a preceding project, the research team developed an initial V0 algorithm for detecting inattentive drivers. This algorithm served as the basis for further development performed during this project.

Event Review Process

In the interest of creating a baseline or ground truth against which to compare algorithm outputs, the research team deemed it necessary to create a reasonable method to assess a baseline level of attentiveness. Ten-second epochs were selected by the research team from which to build a benchmark dataset for algorithm comparisons. Initial events were selected equally across the 24 participants included in the dataset and equally across 0–20 mph, 20–40 mph, 40–60 mph, and 60–80 mph speed bins. After these events were reviewed according to the Distraction Assessment Methodology section, events were added in an attempt to increase the number of events featuring inattentive drivers. These events were sampled from suspected distracted driving cases detected by the V0 algorithm developed in the previous research effort. Through this review, there were a number of events found in which the V0 algorithm incorrectly categorized the driver as inattentive. Actions such as turning at intersections, checking the adjacent lane, etc., were sometimes categorized incorrectly by the original algorithm. These events formed the basis for further algorithm development by providing a set of events where performance of the V0 algorithm should be improved.

Distraction Assessment Methodology

The assessment of distraction levels of the drivers within the events was the key component of the review process. Determining the levels of distraction and how to define them was the initial step of the review process. The table below shows the four levels of distraction that were used in the event review process and how each of the levels was defined by the research team to accurately depict the state of the driver at any given moment. Although subjective in nature, these definitions included key objective indicators (i.e., more heavily distracted drivers generally had a higher proportion of EORT), which facilitated consistent categorization. The categories described below are indicative of our subjective assessments and served as guidelines during the reduction process.

Table 1. Description of Levels of Attention/Inattention

Distraction Level	Descriptor
Not Distracted	Driver is clearly engaged in the driving task, characterized by glances off-road to locations relevant for safe driving
Slightly Distracted	These are cases where driver is looking away from the roadway more than strictly necessary
Moderately Distracted	These cases involve more extended glances off-road, sometimes with phone use or longer uses of the center console
Very Distracted	These cases typically include combined sources of distraction with prolonged glances off-road to a cell phone and the center console

The four graphs in Figure 1 each show DMS output from a different, representative level of distraction. These examples are meant to illustrate typical glance behavior observed across the four distraction levels. "Not Distracted" cases featured primarily on-road glances, for instance, while the opposite is true for events assessed as "Very Distracted."

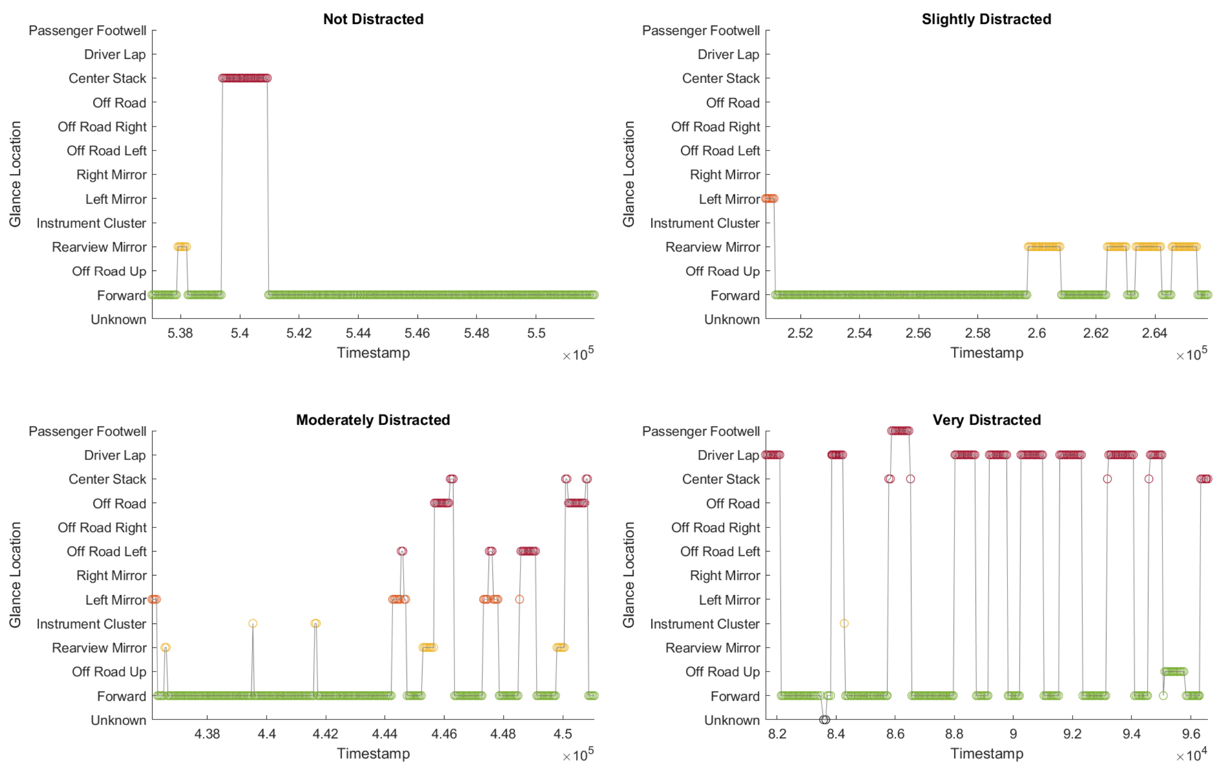

Figure 1. Graphs. Sample data from each level of distraction.

Initially, the same events were reviewed and rated by a pair or researchers independently. Using the definitions from the above Distraction Assessment, the researchers designated a level of distraction observed during the 10-second epoch. This process was completed for a subset of 100 events deemed to be distracted, via the initial V0 algorithm, within the original dataset from which these events were collected. Following the independent review, a calibration review was conducted to ensure that the same approach was being used between both reviewers to ensure consistent categorization of events to the best degree possible. Overall results of the calibration found a 91% agreement level between the two researchers.

Benchmark Event Validation

Per the above approach, the attention levels of the drivers at the end of each event were based on subjective assessments performed by two researchers; in order to directly relate the glance locations to a driver's attention level, a comparison was made between glances off-road and the prescribed attention level. To have enough context for an entire 10-second event (to match the process of the researchers), the glances were aggregated across each event by computing several features using the raw DMS data. These features included the number of glances to each location, duration of each glance, the mean duration of each glance, and the total percentage of each glance location (12) for each 10-second event. In Figure 2, the average percentage of EORT during each event is compared to the driver's attention level. It clearly shows that as a driver's level of distraction increases, the average percentage of time that their glances are off-road also increases. This indicates that our subjective assessments generally aligned with expectations of EORT as it relates to distraction.

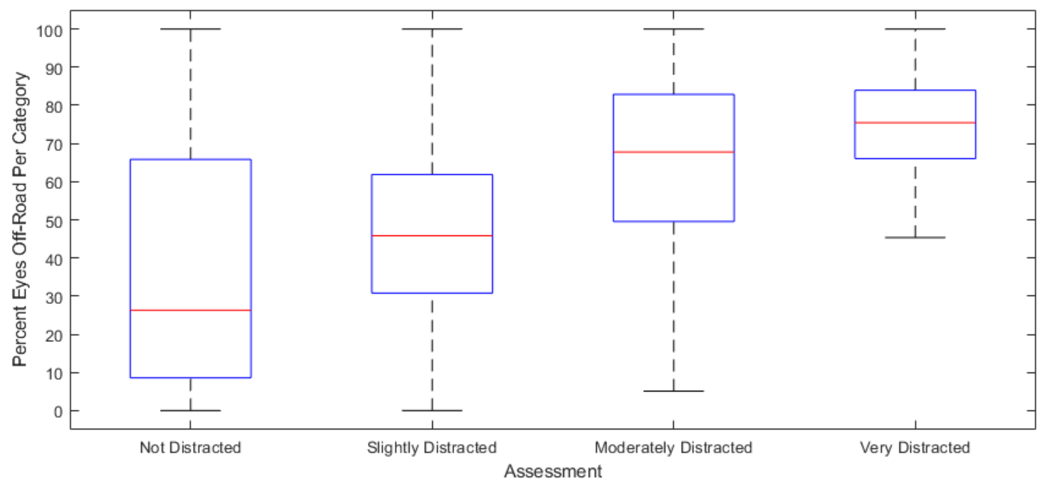

Figure 2. Chart. Mean percentage of EORT for each distraction assessment.

Algorithm Validation Epochs

Additional subsets of epochs were reviewed as the algorithm development continued, and the same process was used to determine drivers' attention status for each event. These epochs were not used to train or modify algorithms, however, and are instead considered a "validation set" to compare algorithm performance between approaches.

DMS Problem Areas

A key area of investigation during the algorithm development phase was to determine areas in which the DMS provided incorrect output, which could impact the attentiveness algorithms. In an attempt to locate problematic areas, a frame-by-frame (FBF) review of accompanying video data for events was carried out by the Virginia Tech Transportation Institute's (VTTI's) data reduction lab. This FBF reduction served as the gold standard, allowing the research team to compare reduction output directly with DMS data. This FBF review identified and recorded the location of the driver's gaze during the selected events, and was then compared directly to the DMS output.

As a result of this review and subsequent comparison, problematic glance locations within the cabin were identified as being more difficult for the DMS to consistently determine the gaze location of the driver.

DMS and FBF Comparison

FBF video reduction was performed on 2,082 10-second epochs (over 600,000 frames of video). Epochs were chosen randomly from the database with the only requirement that the vehicle must be in motion (to eliminate cases where attentiveness was of less consequence due to the vehicle being stopped). Figure 3 provides a confusion matrix for the resulting data, illustrating that while performance is relatively high when drivers are looking forward, the DMS was less reliable in distinguishing between off-road locations that were driving related versus not driving related.

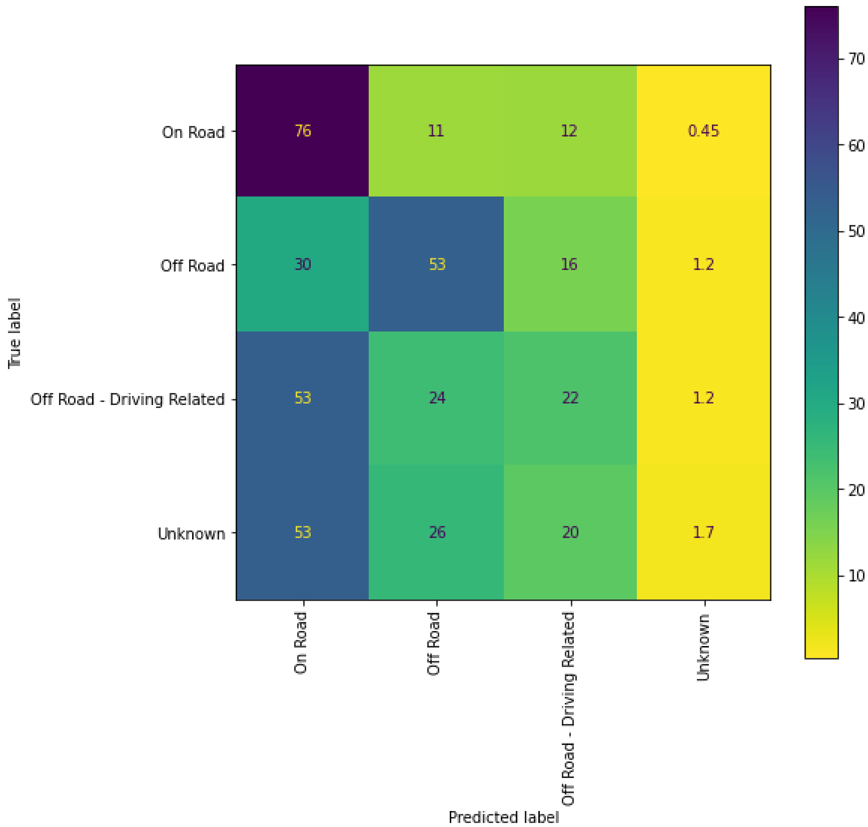

Figure 3. Chart. FBF vs. DMS confusion matrix for glance types.

To better understand the performance limitations of the DMS, epochs with high disagreement between DMS and FBF were identified and manually reviewed by the research team to characterize potential reasons that could explain why the DMS was inaccurate.

Upon review of these events, certain driver behaviors and postures were noted to be problematic and seemed to create errors in gaze location between the DMS and the location identified by the FBF review. Within this subset of events, driver behaviors such as turning, i.e. steering wheel covering the camera view (33%), arm and hand placement (28%), and seating position (5%) were likely impacting the accuracy of the DMS by obscuring the driver's face or removing it from view. In events where driver behavior or posture had no noticeable effect on the accuracy of the DMS, the research team applied further review to determine what other items or behaviors in the driving environment could cause the DMS to be inaccurate. It was noticed that in more than half of these cases, the driver was wearing some form of eyewear, either sunglasses or glasses with corrective lenses. While these findings were not objective or conclusive, they were seen as additional factors that could explain the performance differences between the FBF and DMS assessments.

"Unknown" Events

In addition to reviewing epochs with discrepancies between the DMS and FBF reduction, the research team reviewed instances in which the DMS returned the driver's gaze location as "unknown." This is an extension of investigating problem areas to determine if there are trends within the DMS's characterization of an unknown location or if there were driver behaviors which were contributing to uncertainty on the part of the DMS.

One hundred events were reviewed from both datasets to identify potential causes for unknown classifications. Throughout the subset of events reviewed, a commonality among them was the driver's face being obscured due to driver posture or an object obscuring the DMS, such as the steering wheel or some other object (e.g., a drinking container).

Algorithm Development

The database used for this research project (V0) featured a distraction algorithm developed by VTTI but based in part on methods utilized for the Atten-D algorithm. The algorithm involved increasing and decreasing buffer values based on where the driver was looking at the time. Once the buffers reached a certain value, the driver was determined to either be "attentive" or "inattentive." This general approach formed the basis for the additional exploration in this project. In addition to improving this original approach, two different machine learning methods were pursued that incorporated vehicle data into the algorithmic determinations.

Although numerous variations of algorithms were evaluated, this section focuses on a subset of the notable algorithms.

Buffer-based Algorithm

The buffer-based algorithms created in this study utilize two buffers: AttentionDuration and InAttentionDuration. The AttentionDuration buffer increases when the driver is looking at the forward roadway. The InAttentionDuration buffer increases when the driver is looking away from the forward roadway. Each buffer increases by a factor of the time that the driver is looking at a certain location (in seconds). For instance, if the driver is looking forward for 1 second, then their AttentionDuration would increase by one. Similarly, if the driver is looking to the center stack for 1 second, then their InAttentionDuration would increase by one. Locations such as the instrument cluster or rearview mirror were weighted less heavily, as they are driving related, and so drivers could look to these locations for longer before being marked as InAttentive. Once each buffer increases beyond a threshold, the AttentionStatus (attentive or inattentive) is set, and the other buffer is reset to zero.

Although the initial algorithm only provided an AttentionStatus of either "attentive" or "inattentive," further classifications were created to match the levels of inattention created during the benchmark event review (these classifications are included in algorithm V0). This allowed for a direct comparison between the algorithm output and the attention assessment from the benchmark events, as well as better matching driver behavior observed during the benchmark reduction.

For the initial buffer-based algorithm, glances were sorted into four categories: Attentive, Driving Related (inner), Driving Related (outer), and Inattentive. Each glance type had an associated weighting that impacted its effect on the In/AttentionDuration buffers. Table 2 shows the classification of each glance location and its weighting. The glances that increased the InAttentionDuration buffer are denoted as a negative value. As an example, a 1-second glance to the instrument cluster (with a weighting of −1/3) would increase the InAttentionDuration by only 1/3 of a second. This means that long glances to driving related locations can, correctly, result in a driver being classified as "inattentive."

Table 2. Glance Classification by Location

Glance	Glance Type	Factor
Forward	Attentive	1
Instrument Cluster/Rearview Mirror	Driving Related (inner)	−1/3
Right/Left Window/Mirror	Driving Related (outer)	−2/3
Other	Inattentive	−1

The V0 algorithm resulted in a notable number of false positive determinations of inattentive drivers. Many such events took place in low-speed driving situations where glance patterns deviated significantly from those observed at higher speeds. Figure 5 shows the distribution of mean squared error (MSE) between the algorithm's attention assessment and the actual distraction level of the driver determined during video reduction at low speed (< 25 mph) and high speed (> 25 mph). As seen in the figure, there is a much greater degree of error for the initial algorithm at low speeds than at high speeds. This finding was not surprising given the different glance demands typical between low- and higher-speed driving environments, and indicates a clear area that the research team attempted to improve upon in subsequent algorithms.

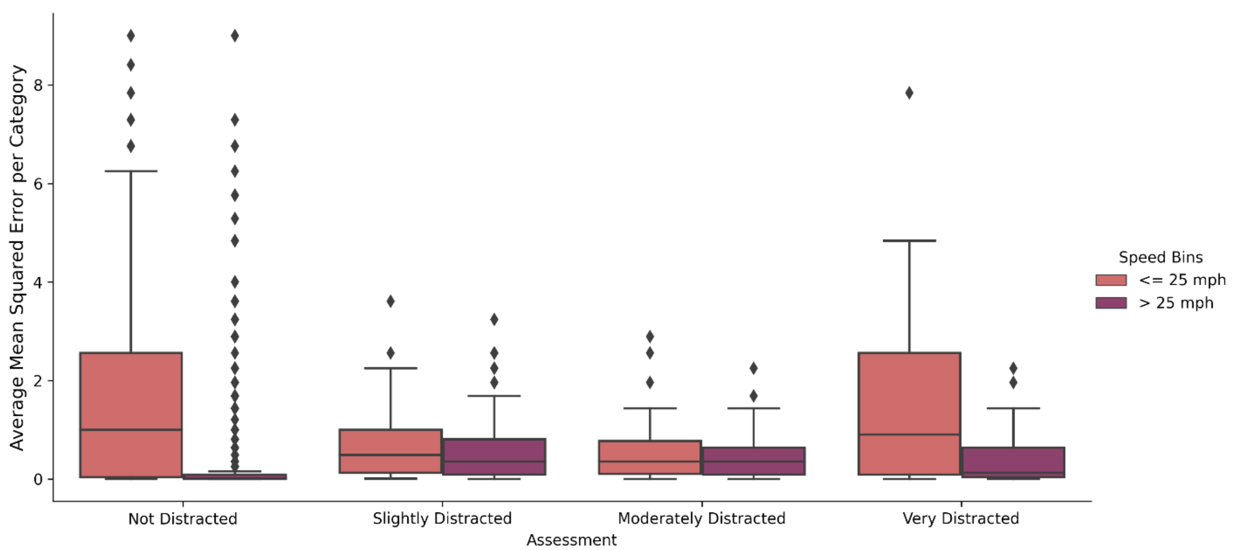

Figure 4. Chart. Initial algorithm error by speed.

V1 Attention Algorithm – Incorporating Speed

The subsequent algorithms developed used the same general methodology as the V0 algorithm, with some minor changes. The speed is incorporated so that the buffers are changed only if the driver is going at least 5 mph. Additionally, the factor by which the buffer is increased is determined by the speed; at set intervals of speed, the InAttentionDuration buffer is increased for inattentive glances, as indicated in Figure 6. A further variation featured a linearly increasing factor between 0 and 65 mph to avoid large jumps in glance weighting as speed moved between thresholds.

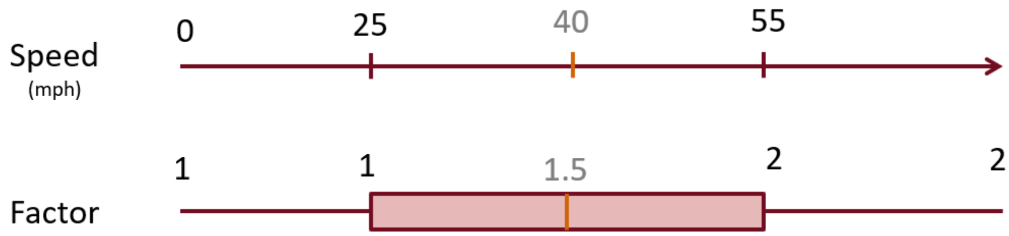

Figure 5. Diagram. V1 speed thresholds.

V6 Attention Algorithm – Parameter Search

The V1 algorithm variation was created by examining trends in the data and carefully crafting thresholds or transformations to affect the classification in a way that increased algorithm performance. Although this did result in improvements compared to the original V0 algorithm, it is slow and, after four additional iterative algorithms it was determined that these methods were too slow for continued exploration.

To counteract this, the team performed a random search on individual parameters to the algorithm. This search allowed for efficient testing of thousands of parameter combinations, something that is simply not possible to manually perform. Parameter combinations were assessed using the average (within each classification type) MSE across the benchmark events, chosen to ensure that larger differences between the expected and predicted classification would result in a lower scoring algorithm. This allowed a comparison of the performance of each individual parameter combination. The individual parameters that the search was performed on, the values that were sampled from, and short descriptions are included in Appendix A – Parameter Search Space.

By including all glance locations individually, theoretically the algorithm can better differentiate between attention types as it is not grouping glance locations and thus is more flexible to patterns observed in the data. Combined, these parameters offer a potential 3,227,550,665,472,000 unique combinations. Obviously, it was unrealistic to sample all combinations within the scope of this effort, and thus a random search was performed with 20,000 combinations randomly chosen for evaluation. The distribution of the average MSE for each combination is shown in Appendix B – Parameter Search Results. The value with the lowest median score was chosen for the final set of parameters as the representative buffer-based algorithm.

Data Bagging

Although the parameter search described in the previous section chose thresholds for classifying driver attentiveness levels, the research team explored using output from the algorithm like InAttentionDuration over a time span to classify driver's attentiveness levels. To do so, we calculated the total maximum of each of these buffers during a driving event, as well as the sum of the local maxima of each buffer during a driving event. Using an ordinal logistic regression model, it was determined that using a combination of these values is best at determining the driver's attention state:

- AttentionDuration and InAttentionDuration at the end of the event
- MaxAttentionDuration and MaxInAttentionDuration during the 10-second event
- SumAttentionDuration and SumInAttentionDuration during the 10-second event

Data bagging takes a sample of a dataset and generates a set of models independently to then apply to an entire dataset with a more accurate prediction. Using a similar technique, we determined the 25^{th}- and 75^{th}-percentile of each of the six values above across all attention states: Not Distracted (0), Slightly Distracted (1), Moderately Distracted (2), and Very Distracted (3). This statistical aggregation then determines which attention state is predicted by each of the six buffer values. Then, an average is taken across all six buffer values to determine the overall attention state of the event (a continuous value between 0 and 3).

Machine Learning Approaches

The more traditional algorithm defined in the previous section is easily interpretable and intuitive to explain, but incorporating additional vehicle information (e.g., steering wheel input) quickly becomes more complex, especially when trying to combine multiple parameters. Due to this limitation, the research team explored two primary machine learning methods for determining driver attentiveness using the available data. The research team's application of machine learning amounts to a classification problem, where data examples are provided to an algorithmic process along with ground-truth values for classification outcome, and the correct classification is "learned" via training. The first method that will be discussed is a neural network approach. The second approach made use of logistic regression.

Neural Networks

As a basis for the neural network architecture used for this classification problem, the team made use of a deep learning model developed by Feng et al. (2021). This model makes use of convolutional layers and gated recurrent units to capture the time-dependent nature of the time-series data used for this project. In the case of Feng et al., they were predicting crashes, near-crashes, and normal driving from acceleration data, but the model was similar enough to prove useful to the attention classification at hand. A diagram showing the full neural network architecture is included below. Outputs were a one-hot encoded value corresponding with one of the four levels of attentiveness established in the Distraction Assessment Methodology section.

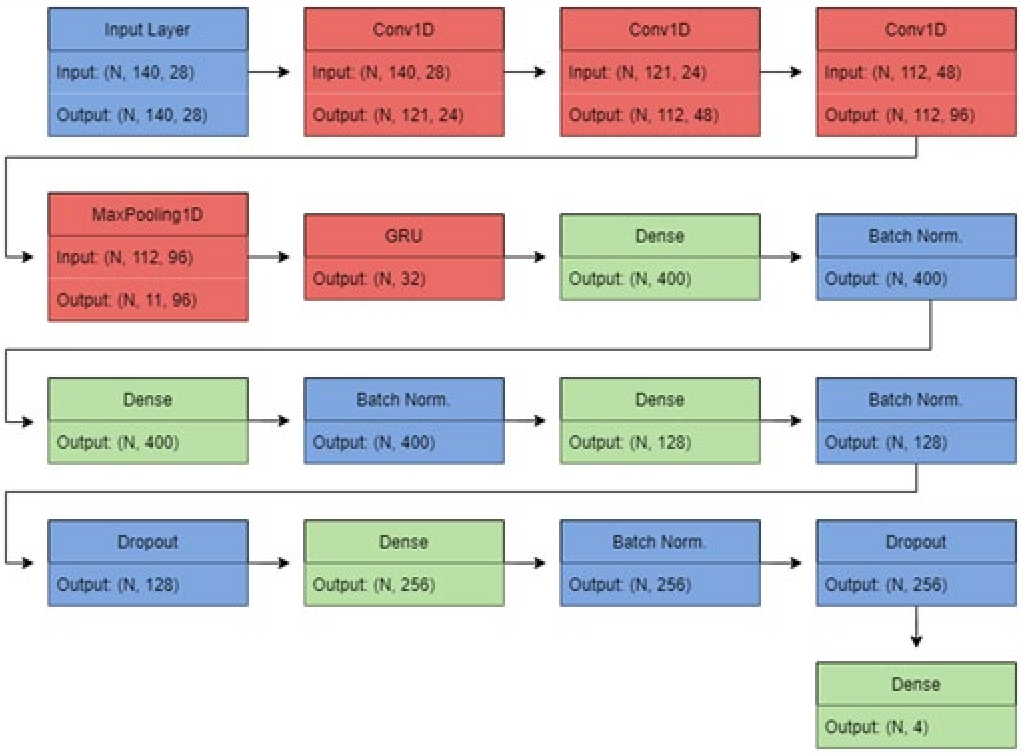

Figure 6. Diagram. Neural network diagram – glance only model.

Prior to training, all data were normalized to values between 0 and 1 using the scikit-learn MinMaxScaler function to improve model learning performance (Pedregosa et al., 2011). All neural network variations were created and trained using the Keras application programming interface (API) (Chollet, 2015) within Tensorflow version 2.4.0 (Abadi et al., 2016). A variation of categorical cross entropy that accounts for ordered classes was used as the loss function (Hart, 2017).

After establishing the base neural network architecture, changes were only made to the input layers of the model to incorporate new sources of data to use in the classification process. Five separate models were created, each trained with a set of data featuring different vehicle parameter data. The different data available to each variation is presented in Table 3.

Table 3. Data Used in Neural Networks

Variation	Glance Location	Speed	Steering Torque	Brake Pedal Position	Throttle Position
NN1	P				
NN2	P	P			
NN3	P	P	P		
NN4	P	P	P	P	P
NN5		P	P	P	P

The more traditional algorithm development showcased the importance of speed in distinguishing between levels of attentiveness, and thus speed is included for each neural network implementation, except for NN1, which served more as a benchmark for performance without additional vehicle measures. Similarly, NN5 also was to serve as a reference for performance using vehicle data alone.

Results

In order to compare the accuracy of each algorithm, the research team developed many different ways to score the algorithms and settled on directly comparing the MSE between the algorithm output and the subjective assessment given to the benchmark events. The attention levels of “Not Distracted,” “Slightly Distracted,” “Moderately Distracted,” and “Very Distracted” were turned into the numerical output of 0, 1, 2, and 3, respectively. Then the events were split up into their benchmark distraction assessments to determine the algorithm accuracy within each category.

Figure 8 shows the algorithm performance versus the benchmark assessment. If the algorithm returned the same AttentionStatus as the benchmark reduction, then it was marked as correct; if it returned a different value then it was marked as some level of false positive (for cases where the algorithm incorrectly marked someone as more inattentive than they were) or false negative (for cases where an algorithm indicated someone was more distracted than they were). If an algorithm was perfect, then there would be green bars across each level of attentiveness. Grades of false positive and false negative are used to indicate whether the algorithm was off by only one level; i.e., slight false positive would apply for cases where the driver was actually “Not Distracted” but the algorithm returned “Slightly Distracted,” and false positive applies when the assessment is off by two or more levels.

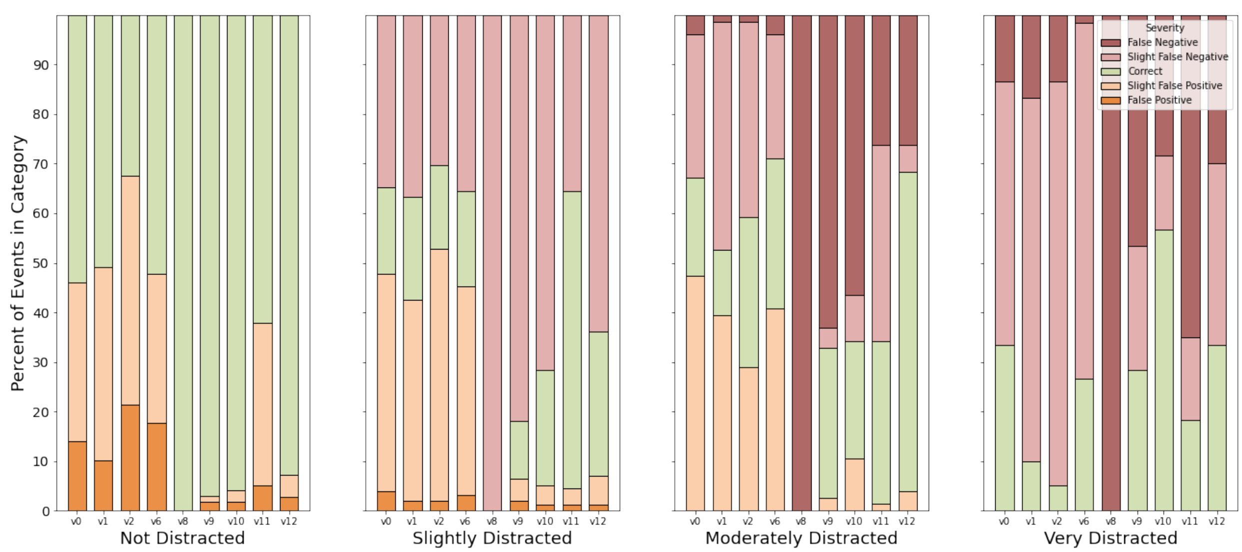

Figure 7. Charts. Algorithm performance by benchmark assessment.

We can see that the buffer-based algorithms (V0, V1, V2, and V6) have consistently fewer false positive cases than their deep learning counterparts (V8–12). The deep learning models do accurately identify more of the moderately distracted and very distracted cases, in general, but also misclassify "Not Distracted" drivers as "Moderately" or "Very Distracted" in a large number of cases. Algorithm V10 is a case study of this, accurately identifying almost 60% of "Very Distracted" cases but also misclassifying ~30% of "Not/Slightly Distracted" cases as "Very Distracted" as well. In general, these algorithms were biased towards classifying drivers as "Not Distracted" regardless of their true AttentionStatus. This theme continues throughout the results, and it is expected that this approach would improve given larger amounts of training data. This sets up an interesting tension between accurately identifying highly distracted cases while keeping the number of misclassifications low, and will be discussed in more detail in coming sections.

Using MSE to compare algorithm performance can identify cases where algorithms perform well in some areas but ultimately underperform due to false negative/positive misclassifications, as was observed for the deep learning examples above. Figure 9 shows the average MSE for each algorithm categorized by the attention assessment determined during video reduction. An ideal algorithm would have 0 error, so lower values indicate higher performance in this chart. This shows that all of the algorithms had a relatively low MSE for the "Not Distracted" events, and that the machine learning algorithms (V8–V12) generally had a larger MSE for more distracted cases. This is indicative of the larger number of false negative classifications in the "Moderately" and "Very Distracted" categories observed in Figure 8.

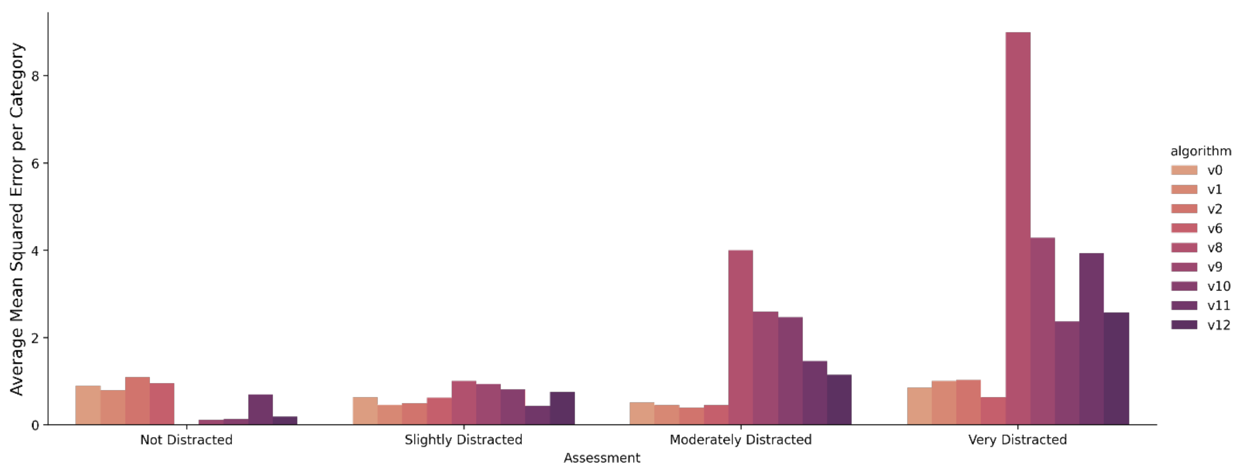

Figure 8. Chart. MSE by benchmark assessment for select algorithms.

Based on the average MSE across each attentiveness level, the research team was able to achieve the greatest overall performance with algorithm V6, a buffer-based algorithm with further processing on the algorithm outputs (AttentionDuration, InAttentionDuration, etc.) over time to determine attentiveness. This algorithm achieved a very high level of accuracy, especially in detecting "Moderately" and "Very Distracted" events. A continuing problem across all buffer-based algorithms, however, is false positive occurrences for real "Not Distracted" cases misclassified as “Moderately Distracted” and “Very Distracted.” Properly assessing driver behavior at all speeds and in all traffic conditions proves very difficult still, but the accuracy observed for true distraction events indicates promise for correctly identifying risky driving behaviors.

Discussion

The research team was consistently challenged by the variation in driver behaviors that fall under the various categories of attentiveness defined in this report. It was discovered quite early that it is exceedingly difficult to use simple heuristics to assign a level of inattention to a driver based on the available data and a human understanding of what distraction looks like. One of the most important discoveries was the differences in glance behavior at high and low speeds. Incorporating this information is integral for distinguishing between attentiveness levels.

DMS accuracy also can be problematic. Although forward glances are typically accurate, distinguishing between other glance locations proves difficult for the systems in use for this study. This leads to uncertainty in algorithm output and accuracy. One way to counteract this would be to bake in probabilistic models based on the accuracy of each location.

Although this study did not directly tackle the question of driver acceptance of distraction detection systems, it is logical to assume that driver response would be more favorable to systems that accurately assess their behavior. Given this assumption, it is critical that accuracy be as high as possible and distraction detection systems reduce false positives. To account for this, the research team used MSE to evaluate algorithm performance, which favored algorithms that have fewer false positive and false negative classifications.

Buffer-based algorithms performed highly in this study. These methods also have the benefit of being interpretable and intuitive for humans to understand. Although this is a benefit to designers and those studying the problem, the type of algorithm deployed is unlikely to have much impact once implemented within vehicles. Still, for improving upon algorithm performance, understanding how an algorithm comes to an assessment is important and can help identify weak points or areas of improvement.

Deep learning approaches presented in this study failed to improve overall performance over even the V0 algorithm, which suffered particularly at low speeds. The research team suspects this is due to the relatively small amount of training data available. Given large amounts of data, deep learning methods are generally effective at identifying patterns and performing classification tasks. Despite a high false positive rate, these models did have a higher classification accuracy for more highly distracted cases, which indicates there is potential benefit in these methods over a buffer-based approach.

An appropriate next step in algorithm assessment would be increasing the amount of available training data and performing algorithm comparisons again. Not only would performance likely improve, but integrating additional data sources into deep learning models is considerably less time-consuming than for the buffer-based models presented in this report. For this reason alone it is important to continue assessing deep learning for applications in this area.

Conclusions and Recommendations

DMSs can be an important component in reducing inattention and crashes related to distraction. The findings presented in this report indicate that both gaze location and vehicle speed at a minimum should be used to assess driver distraction. Speed is particularly important for accurately establishing a driver's attentiveness levels due to large deviations in acceptable glance patterns at low and high speeds. Designing a single algorithm that is effective across a wide range of speeds without incorporating changes based on speed itself is problematic based on our review of the described benchmark events.

It is important to understand that no DMS is likely to be 100% accurate, and therefore algorithms making use of this data should be designed with an understanding of the limitations of this underlying data. If a system is to gain a driver's trust, then it must correspond with their own idea of distraction to an extent. A key component of this is not overestimating the accuracy of the system as to be falsely confident. If a system features an alert component based on driver distraction, but a driver feels they often receive false alerts, then trust is eroded and a driver is more likely to ignore future warnings.

As driver assistance features continue to advance and L2 and L3 ADAS become more commonplace, the driving context and data available to underlying distraction detection algorithms also change. This should have an impact on the design of algorithms for detecting distraction as the level of automation in use by the driver affects the available data for an algorithm to determine attentiveness. For L2 systems, for instance, an algorithm can no longer make use of changes in speed, lane position, and steering wheel torque because the vehicle is now in charge of maintaining the vehicle's positioning and speed. These are all important context to consider as the development of ADAS and DMSs continues into the future.

One last note regarding our approach to detecting driver inattention is worth mentioning: the model we developed did not integrate required scanning of the environment (to maintain situational awareness) into its final assessment. All glances away from the forward roadway were penalized, though driving-related glances were penalized to a lesser extent. For identifying truly attentive drivers, this is potentially insufficient, as scanning mirrors and surrounding traffic is a necessary component of safe driving. Further work in this area, for instance trying to identify drivers who are attentive and ready to resume control of a vehicle with L3 automation capabilities, should likely account for these behaviors. This was considered out of scope for the current research effort as we were more focused on detecting driver distraction rather than levels of readiness or positive attention.

Additional Products

The Education and Workforce Development (EWD) and Technology Transfer (T2) products created as part of this project are described below and are listed on the Safe-D website here. The final project dataset is located on the Safe-D Dataverse.

Education and Workforce Development Products

Through this project we have partially funded a Ph.D. student and contributed to her development as a student and researcher by providing direct experience working with industry data and on a serious problem within the transportation landscape. She has presented the research resulting from this project at campus events and it will form a part of her thesis materials.

Additionally, the content generated from this project is being developed into a learning module for use in the classroom. The project's principal investigator is currently discussing with Virginia Tech faculty the feasibility of presenting this module in their classes.

Technology Transfer Products

Throughout the course of this project, its members have been in communication with an industry champion to discuss results and potential applicability within industry of the algorithm development included in this report. Although it is unlikely that these discussions will result in direct inclusion of these algorithms into vehicles, discussing the ideas and methods used can help demonstrate what is possible in this space.

The student involved in this project has already presented the results of our research at the Lifesavers conference in March 2022. The research team is pursuing a journal publication and plans to support a Safe-D webinar following project conclusion.

Data Products

The data uploaded to the VTTI Dataverse includes 2,082 separate time-series events with driver glance location, speed, braking, and throttle activity. These events can be used as data for training potential attentiveness algorithms. Along with this raw data, output will be included for all of the algorithms discussed in this report. The dataset can be found here.

References

Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, et al. TensorFlow: A System for Large-Scale Machine Learning. *12th USENIX Symposium on Operating Systems Design and Implementation,* (OSDI '16), 2016.

Chung, J., C. Gulcehre, K. Cho, and Y. Bengio. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *NIPS 2014 Deep Learning and Representation Learning Workshops,* 2014.

Collet, F., et al. Keras, 2015. https://keras.io/

Feng, G., C. Qian, L. Shi, and C. Tao. Driving Risk Assessment via High-Frequency, High-Resolution Telematics Data. *Safe-D Webinar.* August 26, 2021.

Hart, J. Keras Ordinal Categorical Crossentropy Loss Function, 2017. https://github.com/JHart96/keras_ordinal_categorical_crossentropy

Kircher, K., and C. Ahlstrom. Evaluation of Methods for the Assessment of Attention While Driving. *Accident Analysis and Prevention,* Vol. 114, 2018, pp. 40-47.

Kircher, K., and C. Ahlstrom. Issues Related to the Driver Distraction Detection Algorithm AttenD. Chalmers University of Technology, 2009.

Klauer, S., T. A. Dingus, and J. D. Sudweeks. The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. NHTSA, U.S. Department of Transportation, 2006.

Liu, A., and D. Salvucci. Modeling and Prediction of Human Driver Behavior. *Intl. Conference on Human-Computer Interaction,* 2001.

Masala, G. L., and E. Grosso. Real Time Detection of Driver Attention: Emerging Solutions Based on Robust Iconic Classifiers and Dictionary of Poses. *Transportation Research Part C: Emerging Technologies,* Vol. 49, 2014, pp. 32-42.

McCall, J. C., and M. M. Trivedi. Visual Context Capture and Analysis for Driver Attention Monitoring. *IEEE Intelligent Transportation Systems Conference,* 2004

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, et al. Scikit-Learn: Machine Learning in Python. *The Journal of Machine Learning Research,* Vol. 12, 2011, pp. 2825-2830.

Smart Eye, and B. Wolfe. Our Eyes or Our Minds – What Causes Driver Distraction? Smart Eye, 2021.

Smith, M. R. H., G. J. Witt, and D. L. Bakowski. A Final Report of Safety Vehicles Using Adaptive Interface Technology: SAVE-IT Summary and Benefits Estimation. NHTSA, U.S. Department of Transportation, 2008.

Tango, F., and M. Botta. Real-Time Detection System of Driver Distraction Using Machine Learning. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 14, Issue 2, 2013, pp. 849-905.

Victor, T., J. Harbluk, and J. A. Engstrom. Sensitivity of Eye-Movement Measures to In-Vehicle Task Difficulty. *Transportation Research Part F: Traffic Psychology and Behaviours,* Vol. 8, 2005, pp. 167-190.

Wang, W., J. Xi, and H. Chen., Modeling and Recognizing Driver Behavior Based on Driving Data: A Survey. *Mathematical Modeling, Analysis, and Advanced Control of Complex Dynamical Systems,* Vol. 2014, 2014.

Wollmer, M., C. Blaschke, B. Schuller, B. Farber, S. Mayer, and B. Trefflich. Online Driver Distraction Detection Using Long Short-Term Memory. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 12, Issue 2, 2011, pp. 574-582.

Appendices

Appendix A – Parameter Search Space

Parameter Name	Description	Values Sampled
Speed Cutoff	Speed (in mph) under which no change to the assessment is made	[0, 5, 10]
Reset Attentiveness Threshold	InAttentionDuration value at which AttentionDuration is reset to 0	[0.5, 0.6, 0.75, 1.0, 0.4]
Reset InAttentiveness Threshold	AttentionDuration value at which InAttentionDuration is reset to 0	[0.5, 0.6, 0.75, 1.0, 0.4]
Slight Threshold	InAttentionDuration value above which the driver becomes "Slightly Distracted"	[1.0, 0.75, 1.25, 1.5]
Moderate Threshold	InAttentionDuration value above which the driver becomes "Moderately Distracted"	[3.0, 1.75, 2.0, 2.5, 4.5]
Very Threshold	InAttentionDuration value above which the driver becomes "Very Distracted"	[5.0, 4.75, 5.25, 5.5]
Speed Threshold	Maximum value at which to linearly scale the speed factor applied to In/AttentionDuration	[15, 25, 35, 45]
Unknown Weight	See note.	[0, -0.5, 0.5]
Forward Weight	See note.	[1, 0.9, 0.8, 1.1, 1.2]
Instrument Cluster Weight	See note.	[-0.33, 0, -0.2, -0.15, 0.1]
Center Stack Weight	See note.	[-1, -0.9, -0.8, -1.1, -1.2, -1.3]
Rearview Mirror Weight	See note.	[-0.33, 0, -0.2, -0.15, 0.1]
Left Mirror Weight	See note.	[-0.5, -0.33, -0.25, -0.6, -0.4, -0.7]
Right Mirror Weight	See note.	[-0.5, -0.33, -0.25, -0.6, -0.4, -0.7]
Driver Lap Weight	See note.	[-1, -0.9, -0.8, -1.1, -1.2, -1.3]
Passenger Footwell Weight	See note.	[-1, -0.9, -0.8, -1.1, -1.2, -1.3]
Off Road Weight	See note.	[-1, -0.9, -0.8, -1.1, -1.2, -1.3]
Off Road Left Weight	See note.	[-1, -0.9, -0.8, -1.1, -1.2, -1.3]
Off Road Right Weight	See note.	[-1, -0.9, -0.8, -1.1, -1.2, -1.3]
Off Road Up Weight	See note.	[1, 0, 0.9, 0.8, -0.5]

Note: Value by which to increase In/AttentionDuration. A value of 1 linearly increases with time, i.e. 1 second looking to a location with a 1 weight would increase the corresponding duration value by 1. A value of 2 would double the increase, and a value of 0.5 would halve the increase. Positive values indicate that AttentionDuration would increase with that weighting, and negative values indicate the same for InAttentionDuration.

Appendix B – Parameter Search Results