Identifying Deviations from Normal Driving Behavior

January 2022 | Final Report

SAFE-D
SAFETY THROUGH DISRUPTION

Virginia Tech Transportation Institute
Virginia Tech

Texas A&M Transportation Institute

San Diego State University
Leadership Starts Here

Disclaimer

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. TTI-Student-08	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Identifying Deviations from Normal Driving Behavior		5. Report Date January, 2022 6. Performing Organization Code:
7. Author(s) Hananeh Alambeigi Anthony D. McDonald Eva Shipp Michael Manser		8. Performing Organization Report No. TTI-Student-08
9. Performing Organization Name and Address: Safe-D National UTC Texas A&M University Texas A&M Transportation Institute 3135 TAMU, College Station, Texas 77843-3135, USA		10. Work Unit No. 11. Contract or Grant No. 69A3551747115/TTI-Student-08
12. Sponsoring Agency Name and Address Office of the Secretary of Transportation (OST) U.S. Department of Transportation (US DOT)		13. Type of Report and Period Final Research Report 14. Sponsoring Agency Code

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.

16. Abstract

One of the critical circumstances in automated vehicle driving is transition of control between partially automated vehicles and drivers. One approach to enhancing the design of transition of control is to predict driver behavior during a takeover by analyzing a driver's state before the takeover occurs. Although there is a wealth of existing driver behavior model prediction literature, little is known regarding takeover performance prediction (e.g., driver error) and its underlying data structure (e.g., window size). Thus, the goal of this study is to predict driver error after a takeover event using supervised machine learning algorithms on various window sizes. Three machine learning algorithms—decision tree, random forest, and support vector machine with a radial basis kernel—were applied to driving performance, physiological, and glance data from a driving simulator experiment examining automated vehicle driving. The results showed that a random forest algorithm with an area under the receiver operating curve of 0.72, trained on a 3 s window before the takeover time, had the highest performance for accurately classifying driver error. In addition, we identified the 10 most critical predictors that resulted in the best error prediction performance. The results of this study can be beneficial for developing driver state algorithms that could be integrated into automated driving systems.

17. Key Words Driver Behavior, Automated Driving, Transfer of Control, Machine Learning Algorithm, Physiological Measures, Predictive Modeling	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.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 22	22. Price $0

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

Abstract

One of the critical circumstances in automated vehicle driving is transition of control between partially automated vehicles and drivers. One approach to enhancing the design of transition of control is to predict driver behavior during a takeover by analyzing a driver's state before the takeover occurs. Although there is a wealth of existing driver behavior model prediction literature, little is known regarding takeover performance prediction (e.g., driver error) and its underlying data structure (e.g., window size). Thus, the goal of this study is to predict driver error after a takeover event using supervised machine learning algorithms on various window sizes. Three machine learning algorithms—decision tree, random forest, and support vector machine with a radial basis kernel—were applied to driving performance, physiological, and glance data from a driving simulator experiment examining automated vehicle driving. The results showed that a random forest algorithm with an area under the receiver operating curve of 0.72, trained on a 3 s window before the takeover time, had the highest performance for accurately classifying driver error. In addition, we identified the 10 most critical predictors that resulted in the best error prediction performance. The results of this study can be beneficial for developing driver state algorithms that could be integrated into automated driving systems.

Acknowledgements

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.

The authors would also like to acknowledge the valuable guidance from John Lenneman of Toyota Motor North America, and Elizabeth Pulver and Scott Christensen from State Farm Mutual Automobile Insurance Company.

Table of Contents

List of Figures

List of Tables

Introduction

Automated vehicle technologies are a promising approach to reduce the nearly 6 million motor vehicle crashes per year in the United States [1]. Yet this promise is limited by the complexity of driver and automated driving system (ADS) interactions [2, 3]. In particular, major challenges regarding safety may arise when drivers are required to take over control of the vehicle after the ADS fails or encounters an operational limit. There is a wealth of existing research that has investigated the influential factors on takeover performance [4]. This research suggests that providing drivers with more time to react to takeovers and assistive technology to aid driver decision making during takeovers may improve takeover performance and safety [4]. Machine learning algorithms that accurately predict post-takeover driver behavior are an important first step in developing such technology. Although a substantial amount of research has been conducted on machine learning in the automated driving domain [4], literature in the area of driver takeover during automated driving is still relatively sparse.

Prior studies have predicted drivers' takeover performance using various machine learning algorithms [5–8]. Du et al. [5] used six machine learning methods to predict the driver's takeover performance, categorizing the performance as bad or good based on the takeover reaction time, maximum resulting acceleration, minimum time to collision (TTC), and standard deviation of road offset. They found that the random forest algorithm on a 3 s time window before the event performed the best when the drivers were engaged with non-driving-related tasks. In another study, Ayoub et al. employed eXtreme Gradient Boosting (XGBoost) to predict the takeover time using variables that influenced takeover time, such as the level of automation and the takeover request modality [6]. That analysis found that the urgency of the situation (low, medium, high), takeover time budget, driver's age, and type of the non-driving-related task (handheld vs. non-handheld) were the most important variables for predicting takeover time. In a study by Braunagel et al., takeover readiness—an indicator of takeover quality—was categorized as low or high and was predicted by three categories of features: the complexity of the traffic situation, the type of secondary task performed by the driver, and the on-road gaze [7]. The study compared support vector machines (SVM) with a linear (SVML) and radial basis kernel, linear discriminant, Naïve Bayes, and k-nearest neighbor (kNN) and found that the SVML had the highest classification performance among the other algorithms. Tivesten et al. [8] developed a simple metric and threshold-based classifier (i.e., a manual approach to select metrics and thresholds that can capture the crash involvement) to predict the driver's takeover performance, categorized as crash and non-crash. Their study analyzed driver glance behavior (e.g., number of on-road and off-road glances) and environmental parameters (e.g., number of issued warnings) and found that a low level of visual attention to the forward road way over a short time window (4 or 6 s), the percentage of time the driver looked off-road for more than 2 or 3 s during the complete drive, and long visual reaction time to attention reminders were associated with increased risk of crash involvement.

In addition to the takeover performance, prior studies have predicted situation awareness [9] and fatigue [10] using various machine learning algorithms. Zhou et al. [9] used Light Gradient Boosting Method (LightGBM) to predict the situation awareness, which was defined as between 0 and 1 based on three performance measures of situation awareness, in recreating simulated driving scenarios during the takeover period. That study used eye-tracking (e.g., number of fixations on the mirrors) and subjective data (e.g., years of driving experiences) as input and found that features such as the length of the video, the time needed to make a decision, and back mirror fixation were the most important in predicting situation awareness over a 1 s time window. Zhou et al. [10] predicted the driver's transition from non-fatigue to fatigue while driving in automated mode using a random forest algorithm and driver physiology over at least a 13.8 s time window (minimum of 1.3 s and maximum of 16.4 s). This analysis found that heart rate, heart rate variability, breathing rate, and standard deviation of breathing rate were the most important features for fatigue prediction. The ground truth, measures, and algorithms used in these studies as well as in the current study are summarized in Table 1.

Although these studies provide valuable insights on driver behavior predictions during automated driving, additional research is needed to understand the features, algorithms, and sampling that best predict driver errors following a takeover. Thus, the goal of this project was to expand the prior analyses to predict driver error during a takeover process using machine learning algorithms on a range of window sizes of a set of driving performance, physiological, and glance data from a driving simulator experiment during partially automated vehicle (SAE Level 2; [11]) driving.

Table 1. The Ground Truth, Measures, and Algorithms from the Literature and the Current Study

Study	Ground Truth	Measures	Algorithm
[5]	Takeover performance (good/bad)	Environmental parameters Physiology Eye glance	SVM NB DA kNN LR RF
[6]	Takeover performance (time)	Environmental parameters Demographics	XGBoost
[7]	Takeover readiness (low/ high)	Environmental parameters Eye glance	Linear SVM SVM with a radial basis kernel NB DA kNN
[8]	Takeover performance (crash/ no-crash)	Environmental parameters Eye glance	Metric and threshold-based
[9]	Situation awareness (0 to 1)	Eye glance	LightGBM
[10]	Fatigue (fatigue/ non-fatigue)	Physiology Eye glance	NARX RF

Current	Takeover performance (post-takeover error/ no-error)	Physiology Eye glance Driving	SVM RBF RF DT

NB: naïve bayes; DA: discriminant analysis; LR: linear regression; RF: random forest; NARX: nonlinear autoregressive exogenous model; RBF: radial basis function; DT: decision tree

Methods

The driving simulation experiment data was collected in Texas A&M Transportation Institute's Realtime Technologies Inc. driving simulator lab. The lab consists of a quarter-cab driving simulator with three screens that provide 165° horizontal and 35° vertical fields of view, a speaker system to provide ambient roadway noise, and a physiological and eye-tracking data collection suite. The driving simulator setup is illustrated in Figure 1. The driving environment and ADS were simulated using SimCreator software and emulated SAE Level 2 automation. The ADS was activated with a button on a touch screen display located to the right of the steering wheel. When the system encountered a failure or an operational limit, the vehicle's ADS was disabled (see [12] for a detailed description).

Figure 1. The driving simulation lab setup including the driver's seat and forward view screens. Note that the eye-tracking system is positioned on top of the dashboard.

Dataset

The study included 64 participants (32 males, 32 females) from the surrounding community, aged 19 to 65, with a mean age of 41.44 (SD = 15.14). All participants were English speakers, reported normal or corrected-to-normal visual acuity and normal color vision, held a valid driver's license, reported driving experience of at least 1.5 years, were not on any medications that may have affected their ability to operate a moving vehicle, had not previously participated in an experiment involving automated vehicles, and had no prior experience driving in an ADS-equipped vehicle (SAE Level 2 and higher). All procedures were approved by the Texas A&M Institutional Review Board (IRB2018-1362D) and were conducted in accordance with the principles expressed in the American Psychological Association Code of Ethics. Informed consent was obtained from each participant and each received $50 for their participation.

Throughout the experiment, the following driving performance data were collected at a 60 Hz sampling rate: continuous steering wheel position, accelerator and brake pedal positions, velocity, time to lane crossing, time headway to an upstream object, and lane position. Physiological indicators, including heart rate, breathing rate, and electrodermal activity (EDA), were also collected from each participant. Heart rate and breathing rate were measured using a Zephyr BioHarness 3.0 (Zephyr Technology, Annapolis, MD, U.S.), an adjustable chest strap, at a 1 Hz sampling rate. The EDA data were measured at 60 Hz using a Shimmer3 wireless Galvanic Skin Response (GSR) sensor (Shimmer, Dublin, Ireland), an elastic strap that was attached to the wrists of subjects' non-dominant hand, and two electrodes attached to the palm. Glance behavior data were collected using a dashboard-mounted FOVIO eye-tracking system (Seeing Machines Inc., Canberra, Australia). The FOVIO system was interfaced with the Eyeworks Data Record software (Eyetracking Inc., Solana Beach, CA, U.S.). Participants were calibrated to the FOVIO system using a four-point calibration screen and were instructed to look at the exterior edges of the panoramic display while maintaining a directly forward field-of-view. The compiled driving performance data set from the experiment is published on the Virginia Tech Transportation Institute's data repository website [12].

The study process consisted of a brief 10 min training on the system's capability and operation, two practice drives, and four counterbalanced experimental drives separated by 2 min breaks. Figure 2 illustrates the temporal depiction of the study process. The experimental drives differed by the type of failure (unexpected braking or obstacle in the road), the deceleration rate of the lead vehicle (2 m/s^2 vs. 5 m/s^2), and the takeover request (alerted vs. silent failure). Regardless of the failure type and criticality, the participants in the silent failure group did not receive any indication of the automation failure. The participants in the alerted group received an auditory and visual alert. The auditory alert consisted of a loud beep, and the visual alert consisted of a change of color on the instrument cluster and an automation activation screen. Only the unexpected braking drives were included in the present analysis because they were deemed most relevant to the project's industry advisors.

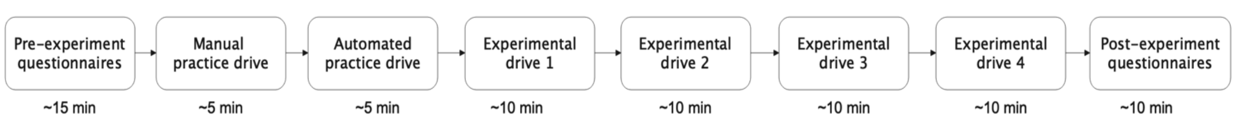

Figure 2. Temporal depiction of the study process.

The experimental drives took place on a 10-mile section of a four-lane straight highway where the participants drove in a three-vehicle platoon with a 1 s time headway. The unexpected braking scenario included one braking event after approximately three and one after seven miles of driving. In the first event, the automation responded to the braking lead vehicle appropriately and in the second event the automation failed to respond. In the latter event, the vehicle's lateral and longitudinal control failed, necessitating a takeover. The criticality of the scenario was manipulated using the deceleration rates of the lead vehicle. Drivers were instructed to keep their hands on the steering wheel throughout the experiment and informed that it was their responsibility to monitor the ADS and the driving environment. Figure 3 shows the unexpected braking takeover scenario from the driver's view (left) and an overhead view (right).

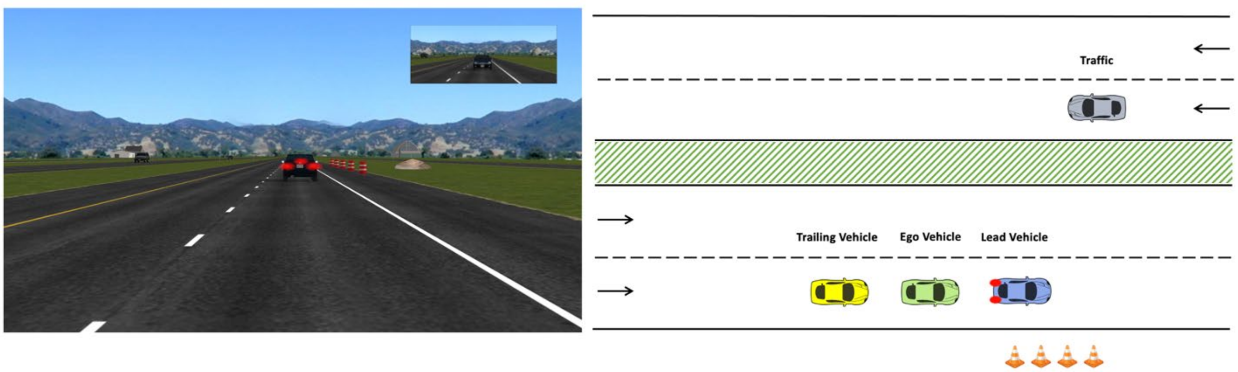

Figure 3. Unexpected braking takeover scenario with the construction zone on the road shoulder. Left: Simulator scenario from the driver's view. Right: Scenario schematic from the top view.

Data Preprocessing and Ground Truth Definition

All 64 participants completed the entire experiment, resulting in 128 completed driving performance, physiological, and glance datasets; however, physiological data from BioHarness—including the heart rate variability—from four drivers were missing and thus excluded. Two additional participants were excluded from the datasets due to eye-tracking technical and calibration issues, resulting in 122 complete datasets. All the data pre-processing and analysis steps were performed in R 4.0.3 [13] using the "tidymodels" package [14]. Figure 4 illustrates the entire analysis schematic.

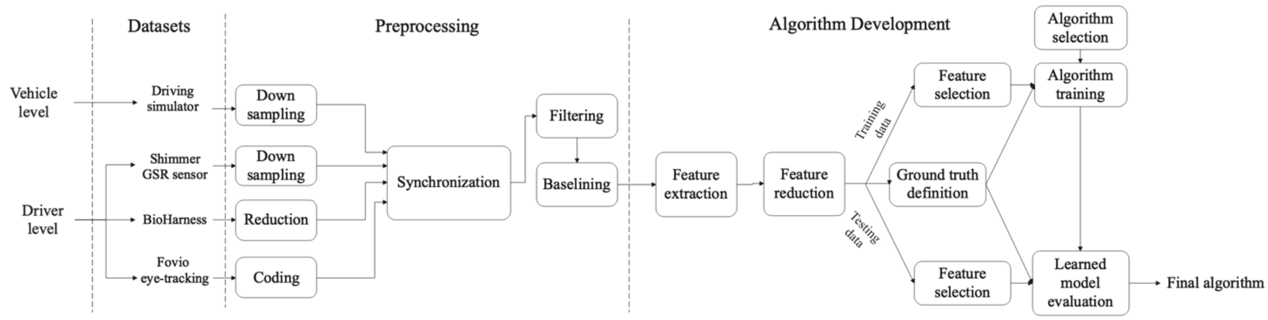

Figure 4. Analysis schematic including the datasets, preprocessing steps, and the algorithm development.

The driving and physiological datasets started from the beginning of the drive and ended approximately 3 min after the second event was completed, where either a crash happened or the situation was resolved. The driving performance data and the physiological data from GSR sensor were down-sampled to 10 Hz from 60 Hz. The data from BioHarness included 1 Hz data from the entire experiment process and was subset into each scenario based on the synchronized time. The glance data were manually annotated by two independent coders from 10 seconds before the event onset until the end of the event. The areas of interest in the coding process included glances at the lead vehicle, dashboard, automation status console, construction site, road, and off-road (e.g., surrounding buildings). For the purpose of the current analysis, the driving performance data only included the time range where the vehicle was manually driven by the driver following the failure. The physiological data included the entire drive (up to the driver's takeover time), and the glance data included 10 s before the event to the time that the takeover maneuver ended. The data from all these sources were time synced to the tenth of a millisecond.

After the data were integrated, a data filtering and baselining process was performed. First, a plausibility filter was applied to the physiological data to remove invalid data (e.g., heart rate values of 0) that were a result of posturing that made the chest strap sensor lose contact with the participant's skin or that were the result of poor fitting that made the chest strap sensor slide against the participant's skin. This step was guided by the data recording limits in each device's user manual. Next, a low-pass Butterworth filter with a sampling frequency of 1 Hz and a cutoff frequency of 0.1 Hz was applied to reduce noise. The optimal cutoff frequency was computed following the work in [15]. Following the noise removal process, the physiological data was scaled relative to a baseline. The baseline was defined as the mean of a 30-second time period of automated driving from the beginning of the drive after the ADS was enabled and before encountering the event for each participant. The selection of this method to define the baseline was guided by prior driving simulator studies [16, 17]. An example of the processed physiological and driving performance data is shown in Figure 5.

The data were labeled as either error or no error, based on the driver's performance during the takeover process using the annotated glance data. Data labeled error were defined as a failure to complete any necessary subtask during the takeover maneuver (e.g., failing to check the side mirror before a lane change) or a failure to complete the necessary tasks in the correct order (e.g., checking the side mirror after a lane change). The accuracy of the error/no error labels was verified by a third independent coder. In total, 22 drives were labeled as error and 100 drives as no error. Table 2 shows the order of subtasks associated with a braking or a lane change maneuver and the categories used to define an error.

Table 2. Order of Subtasks Associated with a Braking or Lane Changing Maneuver and the Categories Used to Define an Error

Maneuver	Subtask	Error
Braking	• Looking at the lead vehicle • Moving hands/feet towards the wheel/pedal • Checking the rear-view mirror • Applying the brake • Avoiding a crash	• Not checking the rear-view mirror • Braking before checking the rear-view mirror • Crash
Lane changing	• Looking at the lead vehicle • Moving hands/feet towards the wheel/pedal • Checking the side-view mirror • Applying the brake • Avoiding a crash	• Not checking the side-view mirror • Lane changing before checking the side-view mirror • Driving off the road • Crash

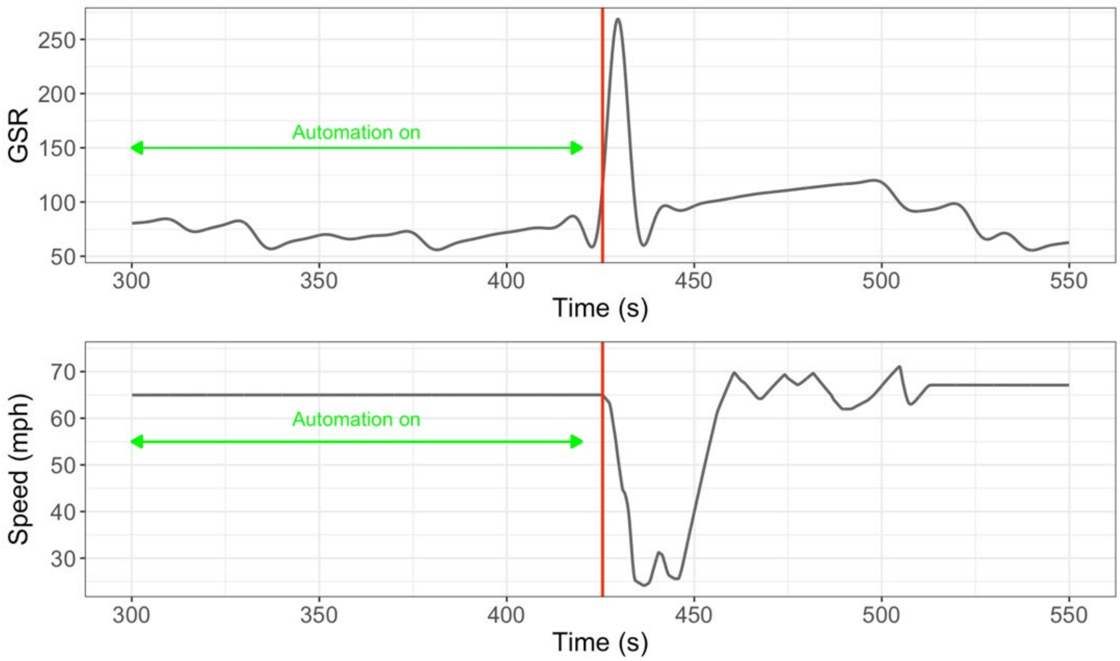

Figure 5. An example of the preprocessed data. The top plot shows the GSR from physiological dataset and the bottom plot shows the speed of the vehicle from the driving simulator dataset. The red line represents the time of the failure.

Feature Extraction and Reduction

Following data preprocessing, a set of 73 features were extracted for window sizes, including 3 s, 5 s, 10 s, 15 s, 20 s, 30 s, 60 s, 120 s, and 300 s before the takeover time. There were no overlaps between the baseline and the features' windows. The driving features were limited to post event whereas the glance features were limited to 10 s before the event. Thus, longer window sizes (> 10 s) mostly consisted of physiological features. The takeover time was defined as the time between the event onset and the start of the maneuver greater than a threshold of 2° for steering wheel angle rotation and 10% for brake pedal position [4]. Features were generated from the driving performance, physiological, and glance measures. Table 3 shows the extracted features along with their corresponding measures and data sources.

Table 3. Categorization of the Datasets, Measures, and the Extracted Features

Data source	Measure	Unit	Feature
Driving simulator	Longitudinal and lateral speed	meter per second	Max, min, mean, med, and std of the speed
Driving simulator	Longitudinal and lateral acceleration	meter per squared second	Max, min, mean, med, and std of the acceleration
Driving simulator	Acceleration and brake pedal position	-	Max, min, mean, med, std, and zero crossing rate of the pedal position
Driving simulator	Lane offset	inch	Max, min, mean, med, std, and lane center crossing rate
Driving simulator	Steering wheel angle	degree	Max, min, mean, med, std, zero crossing rate, maximum steering wheel angle rate, and sample entropy of the steering wheel angle
Driving simulator	Automation disengagement	count	Rate of disengagement
Driving simulator	TTC	second	Min TTC after the event onset
BioHarness/GSR	Heart rate	beats per min	Max, min, mean, med, and std of heart rate
BioHarness/GSR	Heart rate variability	standard deviation in milliseconds	Max, min, mean, med, and std of heart rate variability
BioHarness/GSR	Breathing rate	breaths per minute	Max, min, mean, med, and std of breathing rate
BioHarness/GSR	Galvanic skin response	kilo ohms	Max, min, mean, med, and std of electrodermal activity
Fovio	First fixation location	-	Location of the first observed area of interest after the event onset
Fovio	First fixation duration	second	Duration of the first observed area of interest after the event onset
Fovio	Fixation rate	count	Number of fixations on areas of interests
Fovio	Fixation change rate	count	Number of changes in fixation location
Fovio	Eyes-off-road	second	Duration of off-road glances
Fovio	Eyes-on-road	second	Duration of on-road glances

After the features were generated, they were centered and scaled and feature reduction was performed to remove features with near-zero variance and highly correlated data (features with an absolute Pearson correlation greater than 0.85). To remove the near-zero variance features, two metrics of the percent of unique values (i.e., the number of unique values relative to the total number of samples less than a threshold) and the frequency ratio (i.e., the ratio of the frequency of the most common value to the frequency of the second most common value greater than a threshold) were considered. Following the prior work in the driving performance context, the unique value and the frequency ratio thresholds were set to 10 and 95/5, respectively [18]. Feature removal based on highly correlated features was conducted by analyzing the mean absolute correlation of each highly correlated pair and then removing the feature with the largest mean absolute correlation. This step helped to avoid multicollinearity, which makes the interpretation of the features imprecise and leads to an unreliable prediction. Lastly, the data were up-sampled. These steps were guided by the work in [18]. The feature reduction resulted in a total of 42 features in each window size.

Algorithm Training and Evaluation

Three machine learning algorithms—decision tree, random forest, and SVM with a radial basis kernel—were trained for each of the window size datasets. The training process consisted of a five-fold grouped cross validation process. The data were partitioned at the driver level (to avoid a driver's dataset being included in both training and testing). Following data partitioning, the data were up-sampled to create a training set with an equal amount of error and no error instances. The trained algorithms were assessed by their area under the receiver operating characteristic (ROC) curve (AUC) across the five folds [19], where a higher value of AUC indicates a better performance. The algorithms' AUC differences were statistically evaluated using a bootstrap test (n = 2000) for ROC curves with a threshold of $p < .05$. In addition to the analysis of algorithm performance, a feature importance analysis using permutation-based importance measure was performed for the algorithm with the best predictive performance to provide additional insights into the drivers' behavioral patterns. The feature importance measure was computed by the mean decrease in accuracy.

Results and Discussion

Figure 6 shows the algorithm AUC categorized by the window size and machine learning algorithm (decision tree, random forest, SVM with a radial basis kernel), respectively. The black error bars indicate the 95% confidence intervals. The dashed line in Figure 6 shows random guessing performance. The 3 s time window for the random forest algorithm showed the best performance among the other algorithms and the time windows, where a significant difference was also found in AUC between the random forest with a 3 s time window and random guessing ($p = .01$). Figure 6 shows that the random forest algorithm with a 3 s window size had the highest AUC of 0.72 (0.56, 0.87) followed by a 20 s window with AUC of 0.67 (0.55, 0.78) and 15 s with AUC of 0.65 (0.55, 0.77) for the random forest algorithm. In addition, pairwise comparison showed that this algorithm outperformed random forest across 5, 30, 60, 120, and 300 s ($p < .05$). However, no significant differences were found between 3 s and 10 s ($p = .4$), 15 s ($p = .26$), and 20 s ($p = .5$) windows for the random forest.

The AUC value measures an algorithm's ability to identify errors and distinguish errors from no error cases. Therefore, the results suggest that these algorithms were capable of predicting post-takeover errors significantly better than random guessing, although the findings were inconsistent. Overall, the results showed that the random forest classifier outperformed decision tree and SVM algorithms, as indicated by the AUC values. This finding is consistent with prior studies in the automated vehicle driving domain. Du et al. [5] found that random forest as a classifier had the highest mean prediction accuracy (83%) compared to the other approaches, including decision tree and SVM. In addition, the findings of the current study show that for the random forest model the size of the window significantly influenced the prediction performance. This finding aligns with that of [5], which recommended 3 s as the optimal timeframe to predict drivers' takeover performance. While a broader range of (physiological) measures, up to 300 s before the takeover, were included in this study, no significant improvement was found. Collectively, these findings might suggest that as data gets further from the takeover event (i.e., more than 20 s), the conditions surrounding an error become less prominent or even non-existent. In other words, the correlations between the takeover error and other influential factors might fade, although further investigation is needed to explore this speculation.

In addition to the algorithm analysis, feature importance values were computed to provide additional insight into each feature's relative importance in the post-takeover error prediction. The importance values indicate each feature's mean decrease in accuracy in predicting the error. Thus, larger values of decrease in accuracy represent the most important features. Figure 7 shows the 10 most important features for the random forest model with a 3 s time window. The results showed the importance of all three sources of driving, physiological, and glance measures. The figure indicates that median speed was the most important feature, although features derived from heart rate, glance duration, braking, and steering behavior were also important. This is notable because of the potential implications for data collection requirements for future error-prediction technology.

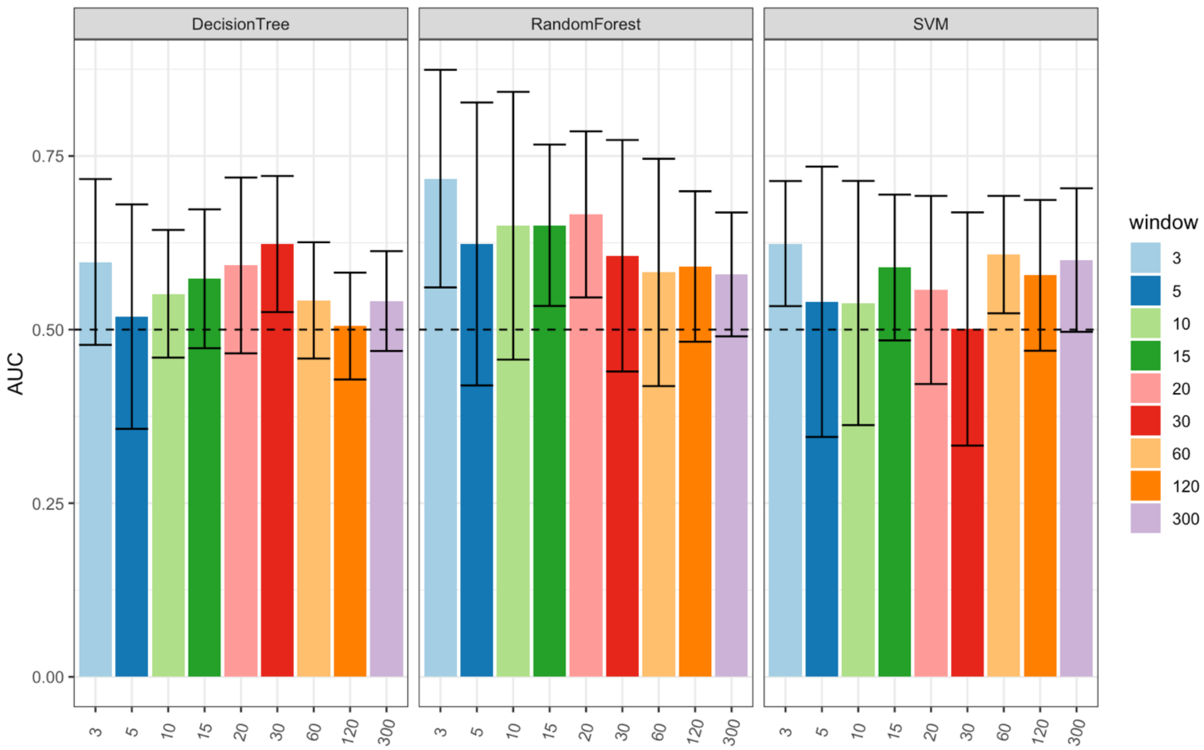

Figure 6. Algorithm AUC categorized by machine learning approach and window size.The horizontal dashed line indicates random guessing performance.

The analysis of feature importance highlights the necessity of including a combination of driving performance, physiological, and glance measures in takeover error prediction. The findings of the 10 most important features show that driving variables, including median lateral and longitudinal speed, median steering wheel angle, minimum TTC, maximum and median brake pedal position, and maximum lateral acceleration, play an important role in error detection. These variables indicate that if the speed, acceleration, or the changes in the steering wheel angle after the failure and prior to the takeover time are higher, this might lead to a more aggressive braking or lane change maneuver. Finding minimum TTC to be one of the most important features is notable. Minimum TTC, which has been used in several studies, is an established surrogate safety metric for longitudinal vehicle control [4] and is defined as the minimum time required to avoid a collision [20]. As a crash gets closer, minimum TTC decreases and therefore inverse TTC, which is a measure of the kinematic severity of a rear-end event [21], increases. In this sense, inverse TTC is associated with the perceived criticality of the situation and has been shown to have a strong link with drivers' behavior, as it may trigger emergency avoidance reactions [22, 23]. Thus, a lower minimum TTC might lead to a more abrupt maneuver and thus more errors. It is notable that these measures are more granular than the environmental parameters (e.g., traffic density) included in prior work [5, 7, 8]. Perhaps the most relevant feature to the findings of this study, in particular to the minimum TTC, is the takeover time budget (i.e., TTC at the time of the failure [4]), that was found in [5] as one of the important features in takeover performance prediction. Prior studies have shown that a shorter takeover time budget is associated with a shorter minimum TTC [24]. Thus, our finding is aligned with the takeover time budget found in [5].

With regards to the physiological indices, heart rate was found to be the most important feature. This result aligns with [5], which also identified heart rate-based measures as important. Duration of on-road glances and first fixation duration 3 s prior to the takeover were the most important glance features. This finding might be associated with the visual readiness component of a takeover process, in which the driver has to redirect gaze to the forward roadway.

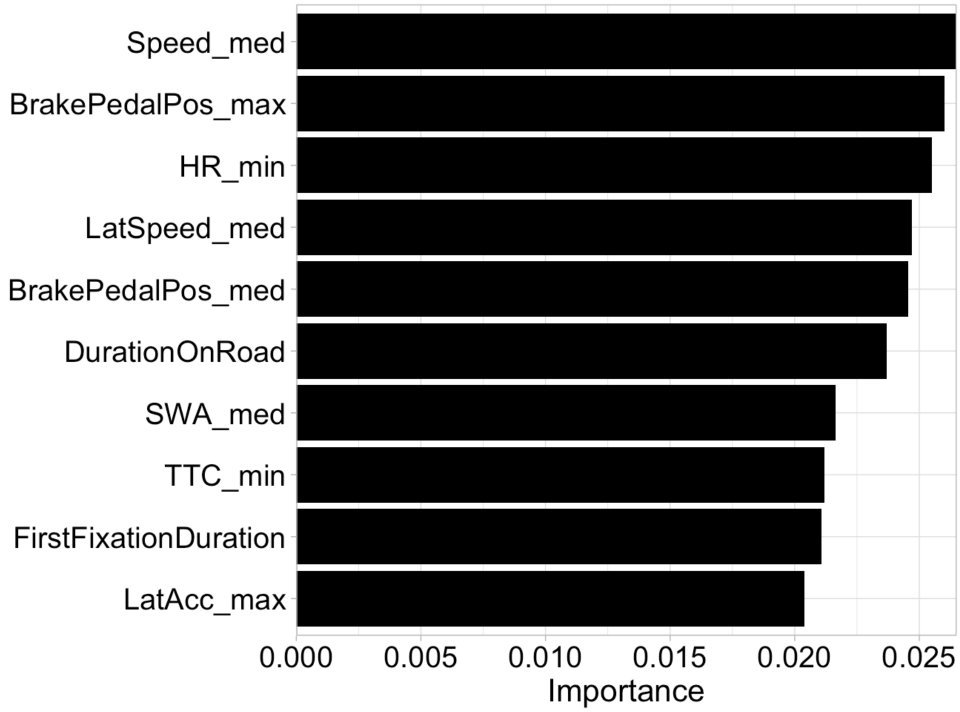

Figure 7. Feature importance values. Med: median; Max: maximum; HR: heart rate; Lat: lateral; SWA: steering wheel angle; TTC: time to collision.

In addition to the previous analysis, to investigate the effects of sampling rate on algorithm performance, a standardized data process was leveraged. Figure 8 shows the results of 1, 5, and 10 Hz sampling rates for three window sizes of 3, 20, and 30 s for the random forest algorithm, which performed the best across all the algorithms. The results show that down-sampling below 10 Hz decreased the performance regardless of window size. The findings suggest a need to further investigate sampling rates above 10 Hz to identify if higher sampling rates will improve performance.

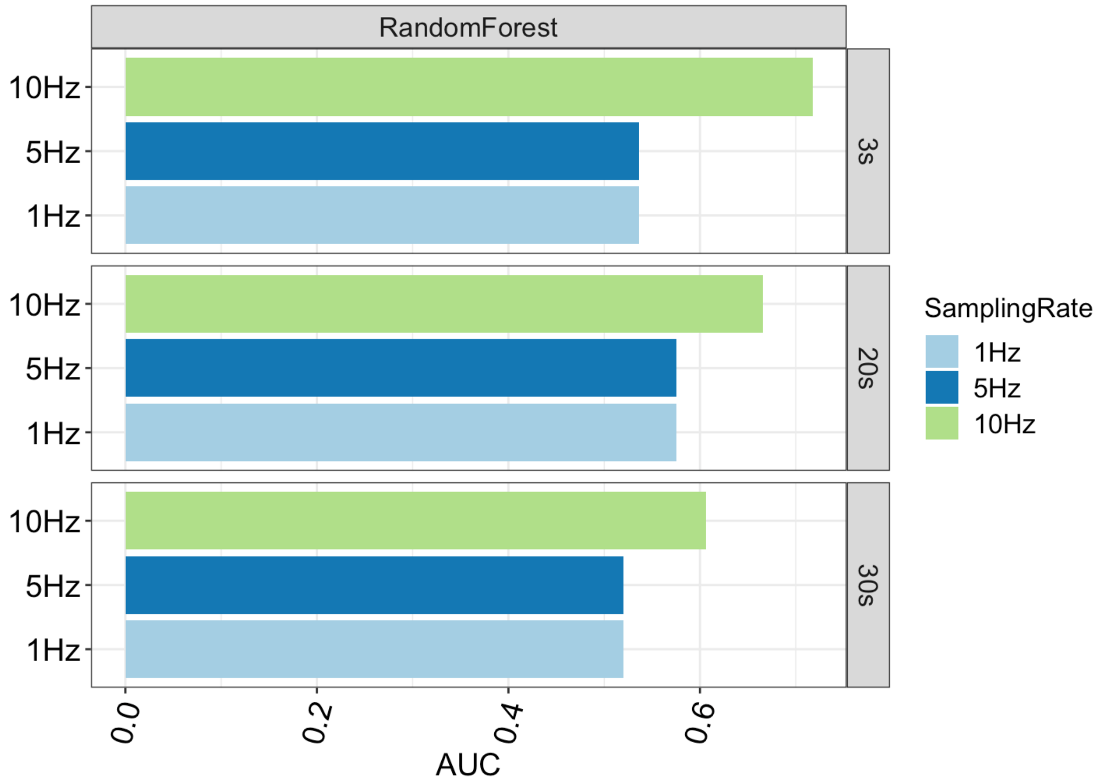

Figure 8. Sampling rate comparison for the random forest algorithm across three window sizes of 3, 20, and 30 s

Limitations and Future Work

Although the analysis provides useful insights with regards to the data implementation and driver error prediction, it is limited in some respects. First, the number of observations in this study was relatively small, which limits the use of more complicated algorithms. Moreover, the collected data were from a driving simulator, which provides relative validity and might not reflect real-world situations. Future work should examine model sensitivity to greater number of variables and data types, investigate whether the validation study results can be extrapolated to a larger pool of drivers, and validate the study in on-road real-world automated driving conditions.

Conclusions and Recommendations

The goal of the current project was to investigate what physiological, driving performance, and glance measures before a takeover can capture driver error during an ADS takeover process. In addition, the study focused on detecting an effective range of data for implementing the model predictions. We analyzed a combination of physiological, driving performance, and glance data from a driving simulator experiment during partially automated vehicle driving where the takeover scenario consisted of a lead vehicle unexpectedly braking due to approaching a construction site. A set of features were generated and three machine learning algorithms—SVM with a radial basis kernel, decision tree, and random forest—were applied to the generated features. We found the random forest with AUC of 0.72 as the best classifier to predict the driver error on the 3 s time window before the takeover time. In addition, the results highlighted the importance of driving performance measures, including speed, brake pedal position, TTC, acceleration and steering wheel angle; physiological measures, including heart rate; and glance measures, including the duration of on-road glances and first fixation duration for predicting driver errors after an ADS failure. The findings provide useful insights for data collection requirements and their implication in designing driver error prediction technologies. In particular, the results have implications in developing algorithms for driver error detection and mitigation. The findings suggest that 3 s prior to the takeover with at least 10 Hz data is required to predict drivers' takeover error. Within this range, driving performance, physiological measures, and glance measures need to be collected. Given these findings, consideration might be given to pursuing additional research that could lead to the development of advanced in-vehicle monitoring systems that proactively monitor the driver's state and issue a warning if an abnormality in the driver's state is detected. Providing dynamic feedback to the driver could potentially mitigate the driver takeover error associated with abnormal driver states.

Additional Products

The Education and Workforce Development (EWD) and Technology Transfer (T2) products created as part of this project can be downloaded from the project page on the Safe-D website. The final project dataset is located on the Safe-D Dataverse.

Education and Workforce Development Products

The project financially supported one Ph.D. student, Hananeh Alambeigi. One undergraduate student, Matthew Buttry, worked on the project as part of a for-credit research course in the Texas A&M Department of Industrial and Systems Engineering and assisted with data preprocessing.

Technology Transfer Products

This project has produced one conference paper and one journal article to date. The conference paper will be presented at the 101st Annual Meeting of the Transportation Research Board [25] and the journal article has been submitted to the Transportation Research Record. In addition, one journal article describing the characteristics of available data sources and variables that can be used in model construction of driver behaviors is planned. We also presented the findings of the project to the National Highway Traffic Safety Administration.

Data Products

The project's used a dataset from a prior SAFE-D project, 03-036. A complete description of the dataset can be found in [12].

References

1. NHTSA's National Center for Statistics and Analysis (2020) Police-Reported Motor Vehicle Traffic Crashes in 2018

2. Lu Z, de Winter JCF (2015) A Review and Framework of Control Authority Transitions in Automated Driving. Procedia Manuf 3:2510–2517 . https://doi.org/10.1016/j.promfg.2015.07.513

3. Alambeigi H, McDonald AD, Tankasala SR (2019) Crash themes in automated vehicles: A topic modeling analysis of the California Department of Motor Vehicles automated vehicle crash database. In: Transportation Research Board 99th Annual Meeting

4. McDonald AD, Alambeigi H, Engström J, Markkula G, Vogelpohl T, Dunne J, Yuma N (2019) Toward Computational Simulations of Behavior During Automated Driving Takeovers: A Review of the Empirical and Modeling Literatures. Hum Factors J Hum Factors Ergon Soc 61:642–688 . https://doi.org/10.1177/0018720819829572

5. Du N, Zhou F, Pulver EM, Tilbury DM, Robert LP, Pradhan AK, Yang XJ (2020) Predicting driver takeover performance in conditionally automated driving. Accid Anal Prev 148:1–11 . https://doi.org/10.1016/j.aap.2020.105748

6. Ayoub J, Du N, Yang XJ, Zhou F (2021) Predicting Driver Takeover Time in Conditionally Automated Driving. XX:1–10

7. Braunagel C, Rosenstiel W, Kasneci E (2017) Ready for take-over? A new driver assistance system for an automated classification of driver take-over readiness. IEEE Intell Transp Syst Mag 9:10–22 . https://doi.org/10.1109/MITS.2017.2743165

8. Tivesten E, Victor TW, Gustavsson P, Johansson J, Aust ML (2019) Out-of-the-loop crash prediction: The automation expectation mismatch (AEM) algorithm. IET Intell Transp Syst 13:1231–1240 . https://doi.org/10.1049/iet-its.2018.5555

9. Zhou F, Yang XJ, de Winter JCF (2021) Using Eye-Tracking Data to Predict Situation Awareness in Real Time During Takeover Transitions in Conditionally Automated Driving. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2021.3069776

10. Zhou F, Alsaid A, Blommer M, Curry R, Swaminathan R, Kochhar D, Talamonti W, Tijerina L, Lei B (2020) Driver fatigue transition prediction in highly automated driving using physiological features. Expert Syst Appl 147:113204 . https://doi.org/10.1016/j.eswa.2020.113204

11. SAE International (2021) Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles

12. Alambeigi H, McDonald AD (2021) Investigating the effects of silent automation failure and scenario criticality on automated vehicle's takeover performance (03-036)

13. R Core Team (2018) R: A language and environment for statistical computing. R Found. Stat. Comput.

14. Kuhn M, Wickham H (2021) Easily Install and Load the "Tidymodels" Packages

15. Yu B, Gabriel D, Noble L, An KN (1999) Estimate of the optimum cutoff frequency for the Butterworth low-pass digital filter. J Appl Biomech 15:318–329 . https://doi.org/10.1123/jab.15.3.318

16. Kim AJ, Alambeigi H, Goddard T, McDonald AD, Anderson BA (2021) Do Threat-Related Attentional Biases Toward Bicyclists Promote or Hinder Safer Driving? Cogn Res Princ Implic Submitted:

17. Son J, Mehler B, Lee T, Park Y, Coughlin J, Reimer B (2011) Impact of Cognitive Workload on Physiological Arousal and Performance in Younger and Older Drivers. In: Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Lake Tahoe, CA., pp 87–94

18. McDonald AD, Ferris TK, Wiener TA (2019) Classification of Driver Distraction: A Comprehensive Analysis of Feature Generation, Machine Learning, and Input Measures. Hum Factors J Hum Factors Ergon Soc 001872081985645 . https://doi.org/10.1177/0018720819856454

19. Fawcett T (2004) ROC graphs: Notes and practical considerations for researchers. Mach Learn 31:1–38

20. SAE International (2013) Operational definitions of driving performance measures and statistics. No. J2944.

21. Markkula G (2014) Driver behavior models for evaluating automotive active safety. Chalmers University of Technology

22. Markkula G, Engström J, Lodin J, Bärgman J, Victor TW (2016) A farewell to brake reaction times? Kinematics-dependent brake response in naturalistic rear-end emergencies. Accid Anal Prev 95:209–226 . https://doi.org/10.1016/j.aap.2016.07.007

23. Engström J (2010) Scenario criticality determines the effects of working memory load on brake response time. In: Krems JF, Petzoldt T, Henning M (eds) Proceedings of the European conference on human centred design for intelligent transport systems (HUMANIST). Lyon, France, pp 25–36

24. Alambeigi H, McDonald AD (2021) A Bayesian Regression Analysis of the Effects of Alert Presence and Scenario Criticality on Automated Vehicle Takeover Performance. Hum Factors J Hum Factors Ergon Soc. https://doi.org/10.1177/00187208211010004

25. Alambeigi H, McDonald AD, Manser M, Shipp E, Lenneman J (2021) Predicting Driver Errors During Automated Vehicle Takeovers. In: Transportation Research Board 101st Annual Meeting (Submitted)