Classification of Driver Distraction: A Comprehensive Analysis of Feature Generation, Machine Learning, and Input Measures

Anthony D. McDonald, **Thomas K. Ferris**, and **Tyler A. Wiener**, Texas A&M University, College Station, USA

Objective: The objective of this study was to analyze a set of driver performance and physiological data using advanced machine learning approaches, including feature generation, to determine the best-performing algorithms for detecting driver distraction and predicting the source of distraction.

Background: Distracted driving is a causal factor in many vehicle crashes, often resulting in injuries and deaths. As mobile devices and in-vehicle information systems become more prevalent, the ability to detect and mitigate driver distraction becomes more important.

Method: This study trained 21 algorithms to identify when drivers were distracted by secondary cognitive and texting tasks. The algorithms included physiological and driving behavioral input processed with a comprehensive feature generation package, Time Series Feature Extraction based on Scalable Hypothesis tests.

Results: Results showed that a Random Forest algorithm, trained using only driving behavior measures and excluding driver physiological data, was the highest-performing algorithm for accurately classifying driver distraction. The most important input measures identified were lane offset, speed, and steering, whereas the most important feature types were standard deviation, quantiles, and nonlinear transforms.

Conclusion: This work suggests that distraction detection algorithms may be improved by considering ensemble machine learning algorithms that are trained with driving behavior measures and nonstandard features. In addition, the study presents several new indicators of distraction derived from speed and steering measures.

Application: Future development of distraction mitigation systems should focus on driver behavior–based algorithms that use complex feature generation techniques.

Keywords: distraction classification, cognitive distraction, machine learning, time-series feature generation, physiological measures

Address correspondence to Anthony D. McDonald, Texas A&M University, 4075 ETB, 3131 TAMU, College Station, TX 77843, USA; e-mail: mcdonald@tamu.edu.

HUMAN FACTORS
Vol. 62, No. 6, September 2020, pp. 1019–1035
DOI: 10.1177/0018720819856454

INTRODUCTION

Driver distraction is a major transportation safety problem. Official analysis of postcrash reports showed that drivers were found to be in a distracted state in at least 14% to 17% of vehicle crashes (National Center for Statistics and Analysis [NCSA], 2017), and other credible estimates show that distraction may be involved in as many as 68% of vehicle crashes (Dingus et al., 2016). The frequency and severity of distraction-affected crashes necessitate a comprehensive solution including legislation, training, and technology.

Although in-vehicle technologies are often a contributing factor in driver distraction, they can also be part of the solution. Distraction mitigation systems can algorithmically combine real-time input from vehicle and driver sensors to estimate the driver distraction, which can then be used to inform adaptive automation such as driver assist systems, alert and guide the driver when attentional reorientation is deemed necessary, and/or provide postdrive feedback to the driver (Kim, Chun, & Dey, 2015; Lee et al., 2013; Schwarz, Brown, Lee, Gaspar, & Kang, 2016; Smith, Witt, Bakowski, Leblanc, & Lee, 2009). Although most prior work in distraction estimation is based on linear or logistic modeling approaches, more recent work has investigated machine learning approaches to facilitate modeling highly nonlinear behavior that can be more sensitive to some types of distraction (e.g., Masood, Rai, Aggarwal, Doja, & Ahmad, 2018; Schwarz et al., 2016).

The challenge of determining when drivers are distracted can be approached using supervised machine learning models. Existing patterns of data associated with different types of distracted driving can be used to develop an algorithm capable of predicting future, unlabeled patterns (Dong, Hu, Uchimura, & Murayama, 2011;

Kotsiantis, 2007). The large body of research on driving distraction has documented how different types of distraction can affect vehicle control input (Dingus et al., 2016; Drews, Yazdani, Godfrey, Cooper, & Strayer, 2009; Engström, Markkula, Victor, & Merat, 2017; Feng et al., 2017; Horrey & Wickens, 2006; Strayer, Drews, & Johnston, 2003), driver head posture and eye-gaze (Lee et al., 2013; Schwarz et al., 2016; Tippey, Sivaraj, & Ferris, 2017), and physiological indicators of arousal in a driver's sympathetic nervous system, such as heart rate measures, galvanic skin response, or perinasal perspiration (Collet, Guillot, & Petit, 2010; Healey & Picard, 2005; Kim et al., 2015; Mehler, Reimer, Coughlin, & Dusek, 2009; Pavlidis et al., 2016; Reimer, Mehler, Coughlin, Roy, & Dusek, 2011). These observable measures can therefore be consulted to provide varying amounts of evidence of a distracted driver. Following this logic, it seems intuitive that an algorithm that learns from multiple vehicle and human variables that are individually sensitive to distraction would improve the sensitivity and specificity of distraction detection. However, few studies of driver distraction have explored machine learning algorithms that include more than one measure (e.g., vehicle and driver physiological measures) in a single algorithm. The studies that have included multiple measures (e.g., Liang & Lee, 2014; Liang, Lee, & Reyes, 2007; Liang, Reyes, & Lee, 2007) have focused on eye-gaze and vehicle control input measures.

Another factor to consider is that different sources of distraction should be mitigated with different solutions. Effective interventions that help a driver recover from distraction due to daydreaming or engaging in a purely cognitive task are not necessarily the best for combating distraction in sensorimotor tasks such as texting on a mobile device (Engström & Victor, 2009). It is therefore an additional challenge to develop "multiclass" machine learning algorithms that can not only detect whether a driver is distracted but provide some insight into the distraction source so that mitigations can be most appropriate for the context.

The current study takes advantage of a large multivariate set of driver behavioral and physiological data collected during human subjects driving simulation studies (Taamneh et al., 2017). Using these data, the performance of several advanced machine learning techniques is investigated toward two goals. The first goal is to develop an effective algorithm that (a) combines driver performance/vehicle input and physiological data sources; (b) is informed by domain knowledge in the generation and selection of complex features; and (c) is designed to distinguish among multiple classes of distraction (nondistraction, cognitive distraction, texting). To the authors' knowledge, such an algorithm has not been introduced in the literature. The second goal of this study is to mine successful predictive algorithms for new insights on driver distraction that can be used to inform mitigation technologies and future experiments. The remainder of this article reviews current detection algorithms, discusses the model training and evaluation process, and discusses the implications of the model findings.

Current Distraction Detection Algorithms

A substantial amount of research has been conducted on supervised machine learning algorithms for distraction detection (Ersal, Fuller, Tsimhoni, Stein, & Fathy, 2010; Li, Jain, & Busso, 2013; Liang & Lee, 2014; Liang, Lee, et al., 2007; Liang, Reyes, et al., 2007; Liu, Yang, Huang, Yeo, & Lin, 2016; Masood et al., 2018; Miyaji, Kawanaka, & Oguri, 2009; Sathyanarayana, Nageswaren, Ghasemzadeh, Jafari, & Hansen, 2008; Son & Park, 2016; Zhang, Owechko, & Zhang, 2004). The algorithms developed by this literature can be characterized by their input data, their ground truth definition of driver distraction, and machine learning approach. Input sources include driving behavior, head and eye tracking, and driver physiological measures. Prior algorithms have used one of two types of ground truth: binary (e.g., distracted cases and normal driving cases) and multiclass (e.g., cognitively distracted cases, texting cases, and normal driving cases). Algorithms that leverage multiple sources of input and multiple classes of ground truth have the most power for inference because supervised machine learning algorithms can only learn from the data and labels provided in their training data set. In this

TABLE 1: Summary of Multiclass Distraction Detection Algorithms

Study	Input Data	Machine Learning Approach	Feature Set	Ground Truth
Torkkola, Massey, and Wood (2004)	Steering angle Accelerator pedal input Lane metrics	Regression Tree	Mean Variance Entropy Stationarity	Multiclass gaze metrics
Li et al. (2013)	Head position Eye closure Speed Steering angle Brake pedal input	k-Nearest Neighbor SVM	Mean Standard deviation Maximum Minimum Range Interquartile range Skew Kurtosis Frequency Duration	Multiclass classifier with cognitive and visual secondary tasks
Son and Park (2016)	Lane position Steering angle	Neural Network	Standard deviation	Multiclass classifier with cognitive and visual secondary tasks
Masood et al. (2018)	Images of the driver and vehicle interior	Convolutional Neural Network	Spatial image features	Multiclass classifier with nine distracting behaviors

Note. SVM = Support Vector Machines.

context, prior work is limited as few approaches have investigated multiclass ground truth. Of the algorithms that have investigated multiclass ground truth, summarized in Table 1, the majority have focused on driver behavior input. The two exceptions, Li et al. (2013) and Masood et al. (2018), augmented driver behavior with head- and eye-tracking measures. Thus, there is a gap in the prior work with multiclass algorithms that use physiological or a combination of driver behavior and physiological measures.

Prior work has explored several machine learning approaches including Support Vector Machines (SVM; Ersal et al., 2010; Jin et al., 2012; Li et al., 2013; Liang, Reyes, et al., 2007; Liu et al., 2016; Miyaji et al., 2009), Bayesian Networks (Liang & Lee, 2014; Liang, Lee, et al., 2007), Neural Networks (NNs; Ersal et al., 2010; Masood et al., 2018; Son & Park, 2016), Decision Trees (DTs; Zhang et al., 2004), Random Forests (RFs; Ragab, Craye, Kamel, & Karray, 2014), and k-Nearest Neighbor (kNN) classifiers (Li et al., 2013; Sathyanarayana et al., 2008). No single approach has shown to significantly outperform all others, although there is some evidence that RFs may outperform DTs and SVM (Ragab et al., 2014). To the best of the authors' knowledge, there has not been a comprehensive comparison of these techniques. In addition, there has not been an attempt to use trained models to find new insights into distracted driving. This study addresses the gaps through a comprehensive comparison of algorithms and a variable importance analysis.

METHODS

Data Collection

The data set used in this analysis was collected in Texas A&M Transportation Institute's (TTI) Realtime Technologies Inc. (RTI) driving simulator (shown in Figure 1). The original goal

Figure 1. The experimental setup including simulator. Note the physiological data collection devices on the participant's wrist.

in the loaded drives were characterized as loading *cognitive* (requiring a verbal response to mathematical or analytical questions), *emotional* (verbal response to emotionally charged questions), or *sensorimotor* (texting on a smartphone provided by the experimenters) resources. A fourth loaded drive included the same driving demands but no secondary task (the *normal* drive). The order of the four loaded drives was randomly counterbalanced across participants. The loaded drive portion of the study was designed as a 2 × 2 × 4 mixed design with Age group and Sex as between-subjects factors, and Load Type (normal, cognitive, emotional, sensorimotor) as a within-subjects factor. The analysis discussed here focused on the cognitive, sensorimotor, and normal loaded drives.

Simulator scenario. The driving scenarios in the four loaded drives were conducted on a 10.9-km section of a four-lane highway with a posted speed limit of 70 km/h, oncoming traffic density of 12 vehicles/km, and 2 buildings/km. Drivers were instructed to drive in the rightmost lane and periodically encountered construction zones in the adjacent lane. Approximately halfway through the drive (5.2 km), drivers were forced to change lanes due to road construction cones blocking the rightmost lane. Participants were instructed to perform secondary tasks (driving the vehicle remained the primary task throughout) during two nonconsecutive phases of the loaded drives. These phases began at 1.2 km into the drive and at 7.2 km into the drive and continued for 3.2 km (see Figure 3). Following the completion of the secondary task, drivers were instructed to resume normal driving. Each loaded drive took approximately 15 min to complete.

Data set. Driving behavior data and driver physiological measures were collected continuously throughout each drive. Driving behavior measures included instantaneous measures of acceleration, brake force, distance, lane offset, lane position, speed, and steering signals. The

of the experiment was to evaluate physiological changes associated with types of distraction (Pavlidis et al., 2016). The experiment began with a total of 78 licensed, regular drivers from two age groups (18–27 years of age; older than 60 years of age); however, 1 participant withdrew due to motion sickness, data from 9 participants were lost due to technical issues, and physiological data were missing for 20 additional participants, resulting in 48 complete data sets. The data sets were approximately balanced across age and gender with 10 males and 14 females in the younger group, and 11 males and 13 females in the older group. The compiled data set from the experiment is published on the Open Science Framework (Taamneh et al., 2017).

Study process. The study included eight experimental phases separated by 2-min breaks, during which participants responded to questionnaires. The eight phases included a relaxation period with no driving task, a practice session on the simulator, a 15-min drive on a 10-km straight road (designed to relax the participants), four drives that included secondary task loading (loaded drives), and a 5-min final drive on a 3.2-km straight road that included a surprise unintended acceleration event. Figure 2 illustrates the full process. The secondary tasks

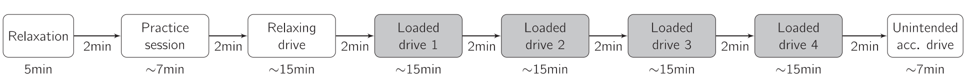

Figure 2. Temporal depiction of the eight experimental phases of the study. The drives included in this analysis are highlighted in gray. ACC = acceleration.

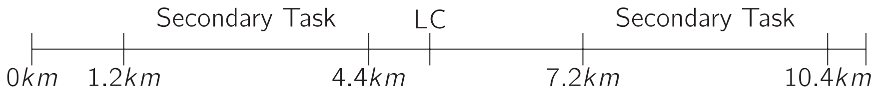

Figure 3. Drive segments by distance into the drive. The secondary task periods represent the portions of the drive where drivers were asked to engage in a secondary task (e.g., texting). LC = the point in the drive where drivers were instructed to perform a lane change.

TABLE 2: Sample of the Raw Data Set Used for Algorithm Training and Evaluations

Time	Secondary Task	Heart Rate	Breathing Rate	Perinasal Perspiration	Speed	Accelerator Pedal Angle	Brake	Steering	Lane Offset
75	0	99	16.2	0.0046	74	5.54	0	0.00	–0.54
76	0	101	16.2	0.0046	74	5.14	0	0.00	–0.50
77	0	99	16.0	0.0047	74	5.50	0	0.00	–0.45
78	0	95	16.0	0.0047	74	5.36	0	–0.01	–0.40
79	0	92	16.2	0.0048	74	4.44	0	–0.01	–0.36
80	1	89	16.2	0.0048	74	0.00	0	–0.01	–0.35
81	1	86	16.1	0.0048	73	0.00	0	0.00	–0.38

Note. Columns including session information (e.g., drive) are omitted for brevity.

driver physiological measures included perinasal perspiration (see Pavlidis et al., 2016), palm electrodermal activity, heart rate, breathing rate, and eye tracking data. The driving and physiological measures were collected at 60 Hz but downsampled to 1 Hz in the published data set. This analysis specifically focused on speed, lane offset, steering angle, brake pedal position, heart rate, perinasal perspiration, and breathing rate, because prior work has illustrated that these measures are sensitive to cognitive distraction, texting, or both. A sample of the full data set is shown in Table 2.

Data Preprocessing

The data set was processed in R (R core team, 2017) through four steps: normalization, windowing, window labeling, and separating into training and testing subsets. The normalization step consisted of subtracting the sample mean from each measure and dividing by the standard deviation. This step was necessary as some machine learning approaches explored here (e.g., SVM) are sensitive to unscaled data. The windowing step involved dividing each drive into nonoverlapping 30-s windows. The window labeling step consisted of assigning a label of cognitive distraction, texting, or normal driving to each 30-s window. Windows were labeled with either the majority class label or, in the case of a tie (which was the case in approximately 7% of the overall data set), the first label in the window was chosen. This method was chosen to account for delays in and lingering effects of the distracting secondary task. A sample of the normalized, windowed, and labeled data set is shown in Table 3. The full data set included 2,060 windows: 470 texting, 504 cognitively distracted, and 1,086 normal.

Finally, the data were randomly split into training and testing data sets, with the training set including 90% of the data and the testing set including 10%. The data were split by driver (rather than by window or by drive) to provide the best possible estimate of how the findings would generalize to an average driver. The testing set was approximately balanced across both age group and sex and contained 231 total windows—53 cognitively distracted, 129 normal driving, and 49 sensorimotor distraction. The training set was further downsampled to achieve an even class distribution. This step was

TABLE 3: Sample of the Normalized, Windowed, and Labeled Data

Window	Time	Train or Test	Label	Normalized HR	Normalized Breathing Rate	Normalized Speed	. . .
1	1	Train	Normal	0.81	0.39	0.38	. . .
1	2	Train	Normal	1.16	0.39	0.38	. . .
1	3	Train	Normal	0.81	0.36	0.38	. . .

Note. Additional columns including other measures are excluded for clarity. HR = heart rate.

TABLE 4: Example of the Reduced Feature Data Set

Window	Train or Test	Label	Lane Offset Standard deviation	Lane Offset Variance	Speed Maximum	. . .
1	Train	Normal	0.448	0.201	0.543	. . .
2	Train	Normal	0.465	0.216	0.035	. . .
3	Train	Normal	0.251	0.063	0.396	. . .

Note. Additional features are not included for clarity.

necessary to avoid bias in algorithm training, for example, algorithms that predominantly predict a single class (Kuhn & Johnson, 2013). After downsampling, the training set contained 421 windows of each class (1,263 total).

Feature Extraction and Reduction

Following data preprocessing, feature extraction and reduction were completed in Python using the Time Series Feature Extraction based on Scalable Hypothesis tests (TSFRESH) package (Christ, Kempa-Liehr, & Feindt, 2016). TSFRESH automatically extracts and filters hundreds of features, including distributional, nonlinear, spectral, Fourier, wavelet, polynomial, and other miscellaneous features such as the frequency of peaks. TSFRESH first calculates all possible features and then performs feature filtering based on a multiple test procedure from the theory of hypothesis testing, which is explained in detail in Christ et al. (2016) and in Benjamini and Yekutieli (2001). The feature reduction step removes both exceptionally rare and exceptionally common feature values from the data. This step reduces the likelihood of identifying coincidental events, such as an animal crossing the road in front of a driver, as distraction events, and improves the speed of algorithm training. After obtaining the output data sets from TSFRESH, additional feature reduction was performed in R. The additional feature reduction included removing features with missing or infinite values, features with zero or near zero variance, and features with high correlations to other features, resulting in a total of 438 features included in each window of the training and testing data sets. A sample of the reduced data set is shown in Table 4.

Algorithm Training and Evaluation

This analysis sought to determine the relative impacts of driver behavioral and physiological features, and of machine learning approaches, on model prediction performance. To accomplish this goal, we analyzed and compared three different data sets for algorithm input:

1. Physiological Data Set: including only driver physiological measures (e.g., breathing rate, heart rate, and perinasal perspiration).
2. Driver Behavior Data Set: including only driving behavior measures (e.g., brake force, lane offset, speed, and steering angle).
3. Combined Data Set: including both driving behavior and driver physiological measures.

In addition to the input sets, seven different machine learning approaches—selected based on the techniques applied in prior literature—were used. The approaches and associated implementation packages are as follows:

1. RF (Liaw & Weiner, 2002)
2. DT (Therneau, Atkinson, & Ripley, 2017)
3. Naïve Bayes (NB; Majka, 2018)
4. kNN (Schliep & Hechenbichler, 2016)
5. SVM with linear kernel function (svmLin; Meyer et al., 2017)
6. SVM with a radial kernel function (svmRad; Karatzoglou, Smola, Hornik, & Zeileis, 2004)
7. NN (Venables & Ripley, 2002)

In total, 21 algorithms were developed, one for each combination of input data type and machine learning approach. All algorithms were trained using the functions in the caret package in R (Kuhn et al., 2017).

RESULTS

The algorithms were compared across their accuracy, average binary class area under the receiver operating characteristic (ROC) curve (Fawcett, 2004), and their confusion matrices for their predictions on the testing data set. Following the algorithm performance analysis, variable importance calculations for the top performing algorithm were used to provide additional insights into behavioral patterns during distracted driving.

Algorithm Classification Performance

The seven machine learning approaches were trained using each of the three input data types and assessed with the testing data set. Figure 4 shows the test set accuracy, along with bootstrapped 95% confidence intervals, and Figure 5 shows the mean test set area under the curve (AUC) for each algorithm, grouped by input data type. In both figures, random guessing performance is shown as a horizontal black line. The SVM algorithm implemented with a linear kernel (i.e., svmLin) is not included in the AUC values because the R implementation provides categorical output and thus thresholding is not possible. The algorithm accuracy differences were statistically evaluated using McNemar's test (Dietterich, 1998) with a threshold of $p < .05$.

Figure 4 illustrates that all but 6 of the 21 algorithms were significantly more accurate

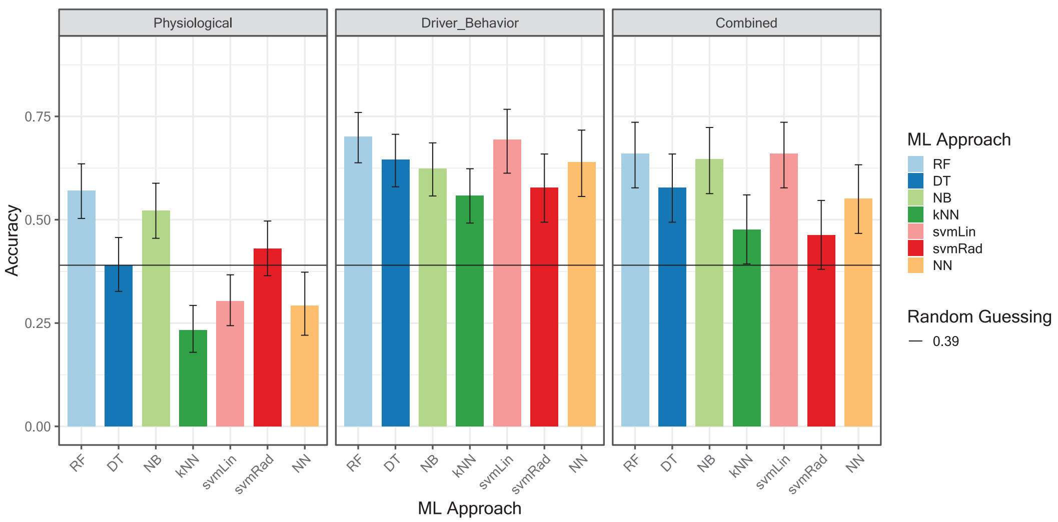

Figure 4. Algorithm accuracy arranged by input measures and ML approach. The black line indicates the accuracy of a random classifier. The error bars represent 95% confidence intervals. ML = machine learning; RF = Random Forest; DT =Decision Tree; NB = Naïve Bayes; kNN = k-Nearest Neighbor; svmLin = Support Vector Machine with linear kernel function; svmRad = Support Vector Machine with a radial kernel function; NN = Neural Network.

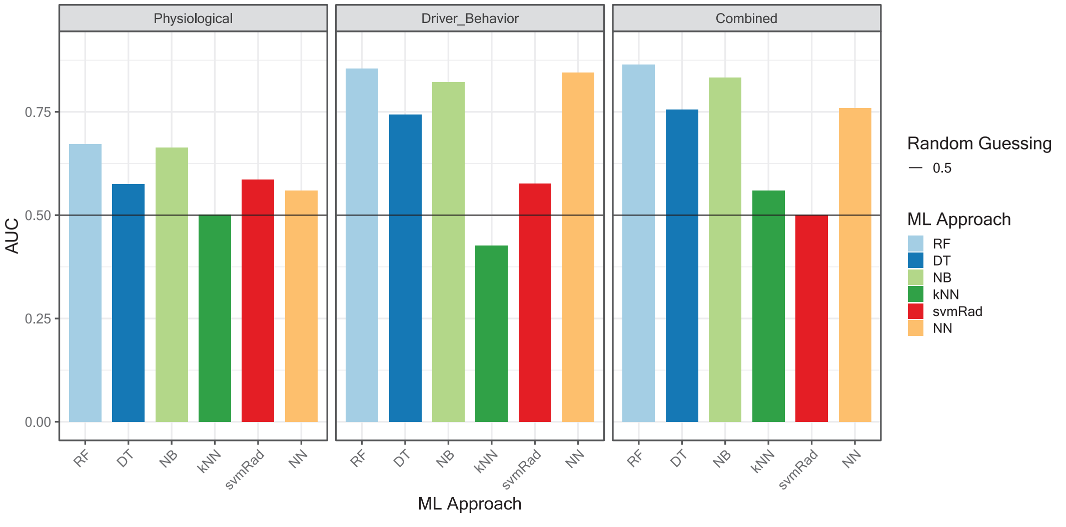

Figure 5. Algorithm average binary AUC arranged by input measures and ML approach. The black line indicates the AUC of a random classifier. Note that confidence intervals are not provided due to the calculation method (Hand & Till, 2001). AUC: area under the ROC curve; ML = machine learning; RF = Random Forest; DT =Decision Tree; NB = Naïve Bayes; kNN = k-Nearest Neighbor; svmRad = Support Vector Machine with a radial kernel function; NN = Neural Network.

than random guessing (all $p < .001$). The 6 algorithms comparable with random guessing used driver physiological data or a combination of physiological data and driver behavior data as input (Physiological DT, kNN, SVM Linear, SVM Radial, and NN, Combined SVM Radial). The RF and SVM with a linear kernel using driver behavior input had the highest accuracy. Pairwise comparisons across the RF models showed that algorithms including driver behavior measures significantly outperformed the physiological algorithm (all $p < .001$). Furthermore, there was no significant difference in accuracy between only driver behavior–based algorithms and the combined algorithms. This result suggests that driving behavior measures dominate physiological measures in the task of classifying driver distraction.

The results in Figures 4 and 5 indicate that some notion of vehicle control is necessary for identifying distraction. More interestingly, the high AUC values in Figure 5 suggest that the driving behavior measures used in this study (i.e., brake force, lane offset, speed, and steering angle) are sufficient to differentiate between external, physical types of distraction such as texting, and internal, mental types of distraction such as solving analytical problems (*cognitive* distraction). A final observation in these figures is that although there were no significant differences between machine learning algorithms, the RF method tended to provide more consistent results across all input types and produced the most accurate algorithm. For this reason and for brevity, the remaining analyses will focus on the RF-based algorithms.

Further insight into the accuracy and AUC results can be gained through analysis of the complete prediction set depicted as a confusion matrix. Figure 6 shows the confusion matrix for the RF algorithms built on Physiological, Driver Behavior, and Combined input data sets. The figure shows the proportion of total training instances (i.e., windows) that were classified correctly (on the diagonal, highlighted in gray) and incorrectly by the algorithms for each class. One unique finding from the figure is that although the physiological algorithm had similar accuracy to random guessing for detecting texting (shown in the bottom right chart of Figure 6), it had comparable accuracy to the other algorithms in detecting cognitive distraction (shown

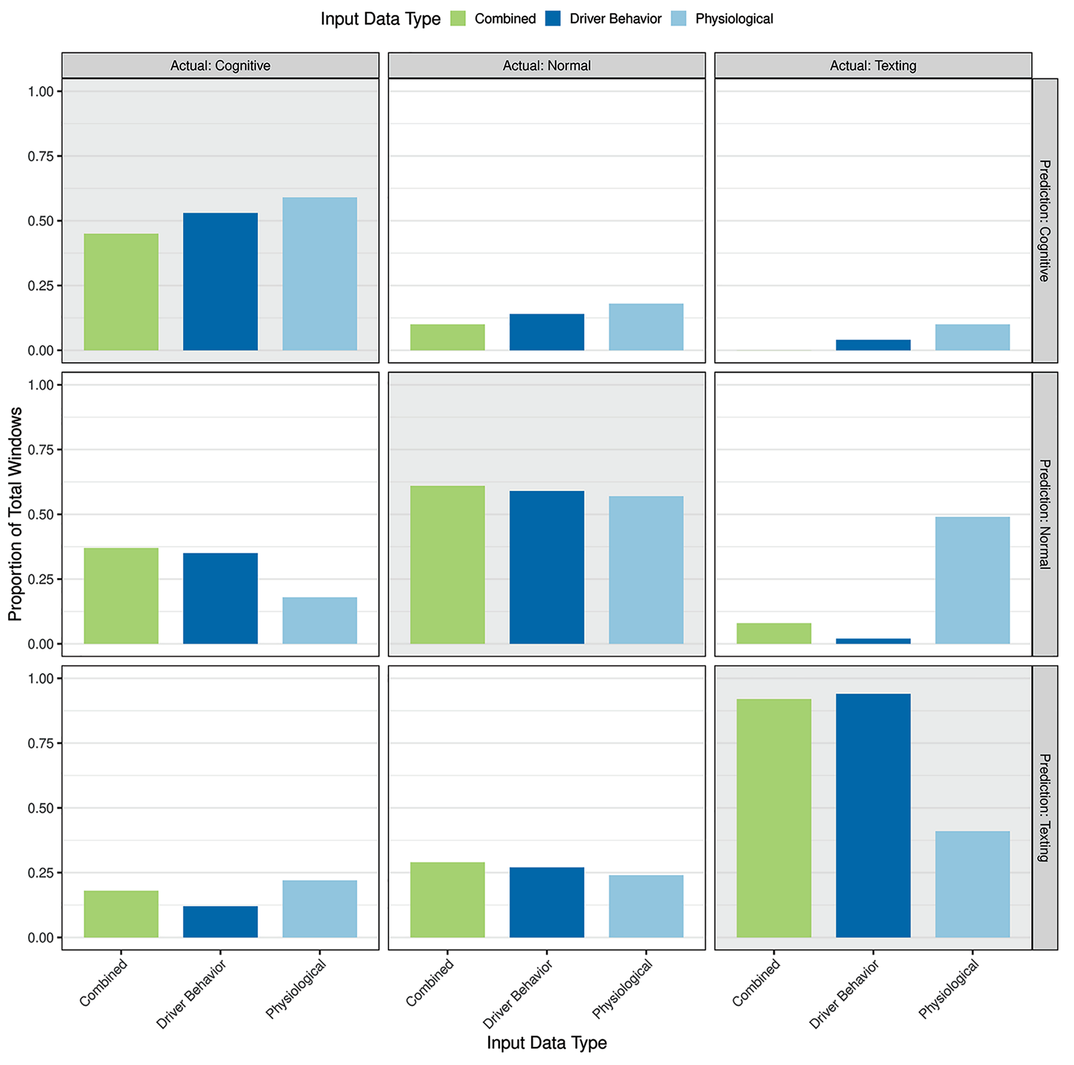

Figure 6. Confusion matrix for Random Forest algorithms and input measures. The gray highlighted charts on the diagonal indicate correct predictions and the off-diagonal plots indicated incorrect predictions.

in the top left chart). In the cases where drivers were actually texting and the algorithms predicted incorrectly (shown in the first and second rows, far right column), the physiological algorithm most often confused the texting cases with normal driving.

Inference With Variable Importance

Although the algorithm performance metrics suggest that driver behavior algorithms can differentiate distraction, they do not support analysis of the relative impacts of features. Variable importance measures, which estimate the mean decrease in accuracy associated with removing a feature from the algorithm, are one way of addressing this gap. Figure 7 shows the 10 most important variables for the RF driver behavior algorithm and Table 5 provides an explanation of each feature.

The results show that lane offset, speed, and steering are the most important measures—that is, they are the most sensitive measures to distraction. Standard deviation of lane offset is the most important feature by a substantial margin, although some nuanced measures including the number of speed measurements greater than the

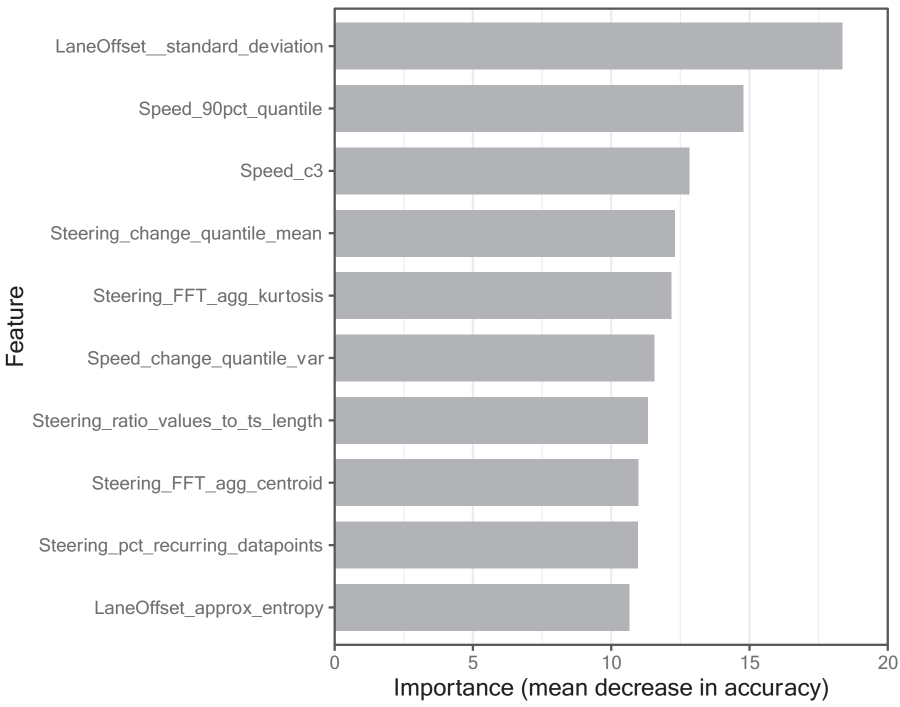

Figure 7. Variable importance values—measured by mean decrease in accuracy associated with removing the feature—for the 10 most important features in the Random Forest driver behavior algorithm. FFT = Fast Fourier Transform.

90th percentile quantile, the C3 measure of speed, and the change quantiles of steering also have substantial impact on algorithm performance. These features are, for the most part, not represented in prior analyses of distraction behavior including distraction algorithm development.

Beyond these findings, further insight can be achieved by analyzing the distribution of features across classes. Although analyzing these distributions in isolation negates the power of machine learning methods to find complex patterns across features, it provides some insight to the patterns in the data associated with classes. Figure 8 shows an example of this analysis—a violin plot—partitioned by feature and distraction type and ordered by variable importance. Each violin plot shows the distribution a feature mirrored across the horizontal axis (Hintze & Nelson, 1998). Differences in the means (shown as points on the plot) or the shapes of the distributions between the classes suggest that the measures may be useful indicators of distraction.

The plots show clear differences in the distributions of data associated with cognitive distraction, normal driving, and texting in the four most important variables (i.e., the top four charts in the figure). In particular, the standard deviation of lane offset chart shows that texting drivers had an almost uniform distribution, whereas cognitively distracted and normal drivers were more normally distributed. The mean standard deviation of lane offset is lowest in the cognitively distracted cases and highest in the texting cases. The 90th percentile speed plot shows a lower mean value for texting compared with normal driving and cognitive distraction and a broader variance. Thus, texting drivers, on average, have fewer instances of speeds greater than the 90th percentile than normal drivers. In both cases, the differences in the distributions suggest that these metrics would be viable indicators of distraction, even for analyses that do not use

TABLE 5: Description of the Most Important Features of the Random Forest Driver Behavior Algorithm

Measure	Feature Type	Description
Lane offset	Standard deviation	Standard deviation of the time series.
Speed	90% quantile	The number of data points greater than the 90% quantile of the time series.
Speed	C3	A measure of nonlinearity of a time series calculated as the autocovariance of the current and two previous values of the time series (Schreiber & Schmitz, 1997).
Steering	Change quantile mean	The average absolute value of the changes in the time series.
Steering	FFT aggregated kurtosis	The kurtosis of the FFT of the time series.
Speed	Change quantile variance	The variance of the absolute value of the consecutive changes in the time series.
Steering	Ratio value to time-series length	The ratio of the number of unique values to the total length of the time series.
Steering	FFT aggregated centroid	The mean frequency of the FFT of the time series.
Steering	Percentage of recurring data points	The percentage of unique values in the time series. The value is 1 if all values are unique and less than 1 if multiple values are repeated.
Lane offset	Approximate entropy	A measure of the level of randomness of the time series.

Note. FFT = Fast Fourier Transform.

machine learning, such as analysis of variance (ANOVA). As the variable importance decreases (e.g., in the lane offset approximate entropy variable), there is more overlap in the distributions and thus it is less likely that these variables would show significant differences in empirical analyses using linear models. However, pairwise Kolmogorov–Smirnov tests—which assess whether samples of data originate from different distributions—showed that all of the distributions shown in Figure 8 are significantly different (all $p < .001$). Collectively, these results indicate that there may be a benefit to using 90th percentile quantiles, nonlinearity (C3), and change quantile metrics as dependent measures of distraction in subsequent analyses. Such measures may be more sensitive to distraction than other more common metrics such as mean speed and steering reversal rates, which were also included in this analysis but found to be less important for classification.

DISCUSSION

Distracted driving remains one of the largest safety risks to current ground transportation systems (Dingus et al., 2016; NCSA, 2017). Although past work has advanced the power of distraction detection algorithms, few attempt multiclass classification (e.g., Li et al., 2013), and fewer systematically compare advanced machine learning techniques. The classification aspect is important for effectively mitigating distraction, as cognitive distraction and texting-related distraction should be mitigated with different strategies. Very few studies published in driving contexts have compared and contrasted input data sets that included physiological, behavioral, or both types of data, and to the authors' knowledge, none have attempted this comparison with the inclusion of complex, nonlinear features that may be maximally sensitive to certain types of distraction. The results of this work can inform human performance modeling and driver distraction algorithm development and illustrate the value of advanced machine learning techniques for inferring human states.

Algorithm Classification Performance

The performance of distraction classification algorithms was affected more by the selected

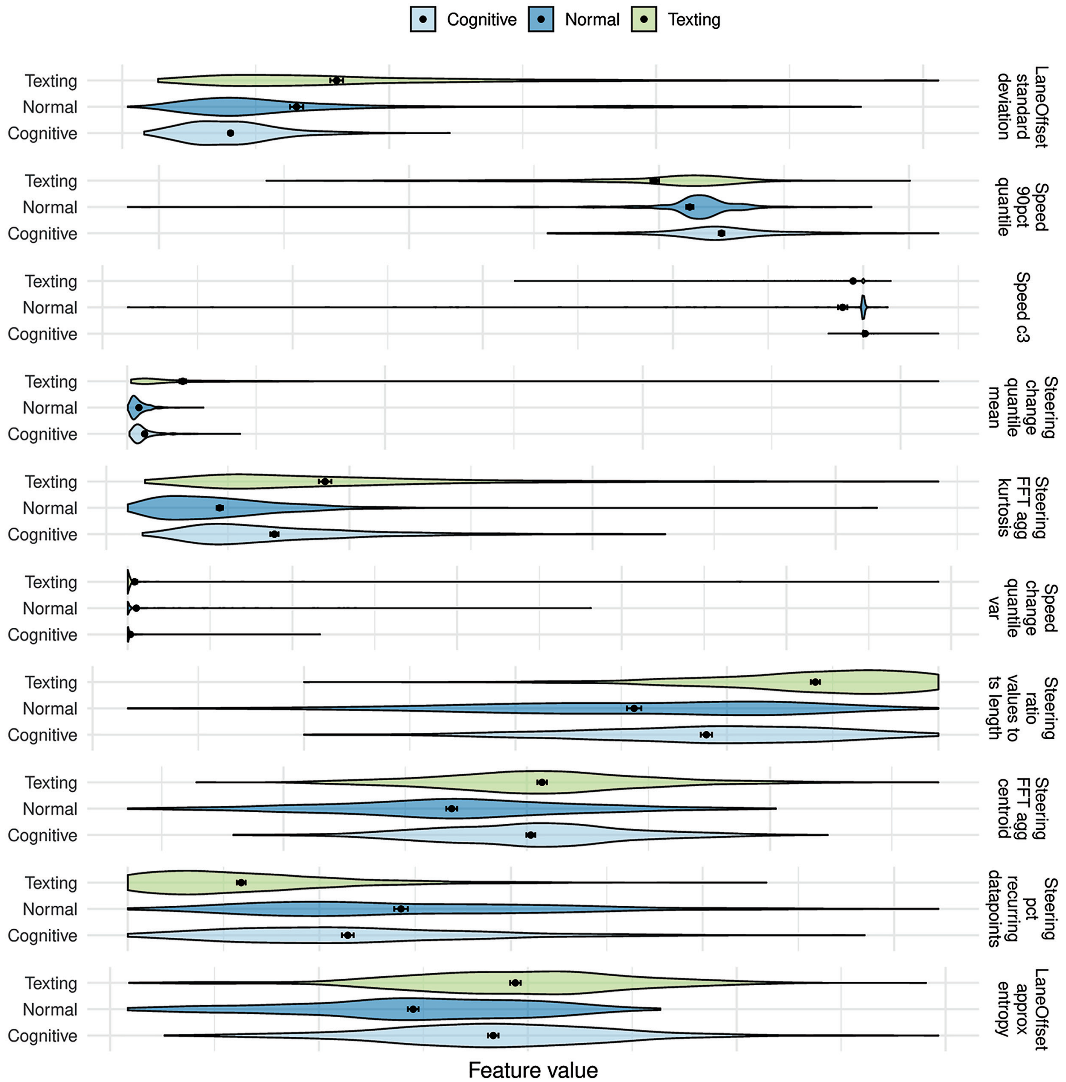

Figure 8. Violin plots of feature values by distraction condition. FFT = Fast Fourier Transform.

input measures (e.g., driving performance and/or physiological data, as well as generated features) than by the machine learning approaches explored here. The accuracy and AUC results show that driver behavior (or a combination of driver behavior and physiological) measures are effective for differentiating cognitive distraction and texting from normal driving. Algorithms that use physiological measures alone were able to differentiate normal driving from states of cognitive distraction, but not from sensorimotor distraction (texting). This is a noteworthy limitation, as distracting activities that engage sensorimotor resources are associated with some of the largest increases in crash risk (Caird, Johnston, Willness, Asbridge, & Steel, 2014; Caird, Simmons, Wiley, Johnston, & Horrey, 2018; Klauer et al., 2014) and thus should be emphasized in the development of distraction detection and mitigation systems.

The findings from the current study add to a mixed body of evidence from research seeking to infer operator cognitive states by combining and comparing various performance-based, subjective, and physiological measures. In a driving simulator, Hicks and Wierwille (1979) compared subjective measures, driver behavioral data, and physiological correlates of workload. Similar to the results of the current study, driving measures—specifically, steering and lateral deviation metrics—were found to be considerably more sensitive than physiological measures to the effects of imposed workload. Similarly, Engström, Johansson, and Östlund (2005) and Pavlidis et al. (2016) showed that when task-imposed workload is increased in a driving simulation, lateral and longitudinal control measures are reliably sensitive to changes in workload, whereas physiological measures are comparably less sensitive.

In contrast, other research has found physiological measures to be highly sensitive to changes in stress or imposed task load on drivers (Brookhuis & de Waard, 2010). Healey and Picard (2005) showed that sophisticated algorithms based on physiological measures can distinguish up to three levels of driver stress with accuracies as high as 97%. Mehler, Reimer, and Coughlin (2012) found greater sensitivities for physiological measures than for driving-performance measures when discriminating among levels of cognitive demand. Similarly, Solovey, Zec, Perez, Reimer, and Mehler (2014) found an improvement in workload discrimination when physiological data were included in machine learning algorithms along with driver performance data.

The inconsistency in historical results comparing physiological and performance-based indicators of driver cognitive state may be partially due to methodological factors such as participant selection, training methods, task selection and pacing, and the scope of allowable driver behaviors (Mehler et al., 2012). The inconsistency may also reflect the relationship between imposed task loads and one's personal cognitive load limit, sometimes referred to as the "redline" of cognitive workload (Grier et al., 2008; Wickens, Hollands, Banbury, & Parasuraman, 2013). Although task loads are below one's redline, performance is often only minimally affected by changing load levels, especially when tasks are characterized by high levels of automaticity (Engström et al., 2017), because drivers have resources available to engage as more support is needed (Mehler et al., 2012). At these "sub-redline" workload levels, driving-performance measures are therefore less sensitive than physiological indicators of workload. This logic can be applied to the development of physiological algorithms that can detect and trigger mitigation strategies when operator workloads approach the redline (Rodriguez-Paras, Susindar, Lee, & Ferris, 2017; Rodriguez-Paras, Yang, Tippey, & Ferris, 2015) to minimize performance and safety decrements. Indeed, the accuracy and AUC results from this study suggest that algorithms that combine driver behavior and physiological changes may be robust to this redline performance; however, the results should be confirmed by a more rigorous analysis.

Inferential Analysis

The variable importance analysis found that lane offset, speed, and steering provide the most sensitive measures in the distraction classification algorithm. Lane offset standard deviation is the most sensitive metric, followed by the frequency of speeds greater than the 90% quantile, the nonlinearity of the speed (C3), the mean of the absolute value of changes in steering (change entropy), and the aggregated kurtosis of the Fast Fourier Transform (FFT) of the steering signal. These findings align with prior work that has indicated that both cognitive and texting distractions affected lane position variance (Cooper, Medeiros-Ward, & Strayer, 2013; Drews et al., 2009; Engström et al., 2005, 2017; Liang & Lee, 2010). Although the current study included simple steering metrics commonly used to assess distraction, such as steering reversal rate (Macdonald & Hoffmann, 1980) or steering entropy (Boer, 2000), the feature importance analysis showed that more complex, nonlinear transformations of steering data, such as steering change quantiles and FFT metrics, were more sensitive to the types of distraction investigated in this work.

These findings are noteworthy because they indicate how nonlinear complex features may offer a clear direction for improving on existing driver distraction algorithms and analyses built on simpler metrics. For example, there has been considerable divergence in how measures of speed (as well as other performance measures) are associated with different types of distraction (e.g., Engström et al., 2005, 2017). The findings here offer the possibility that some of this divergence is related to the manner in which speed is represented in models. Although most prior studies have examined linear patterns in mean speed, the current study found speed measures above the 90% quantile to be far more sensitive and specific for classifying states of distraction. Further research can investigate whether such advanced metrics may be more robust to individual differences in driving styles, which represent a key challenge in this research (Engström et al., 2005).

The algorithms in the current work were designed to distinguish three driver states which included normal driving and driving while distracted by either cognitive or sensorimotor-based secondary tasks. Most prior work focuses on unidimensional assessments that are primarily quantitative, designed to support additive comparisons in workload levels (e.g., "high" vs. "low" levels). In contrast, the current classification effort attempts to qualitatively infer the nature of the distraction by comparing measures that respond differently when loading cognitive, physical, and visual resources (e.g., Brookhuis & de Waard, 2010; Engström et al., 2005). Future research may explore a wider set of distracting tasks with special attention given to machine learning input features that are differentially sensitive to different types of loading.

Limitations and Future Work

The limitations of this analysis include the simulator and study design, the instrumentation and sensors used, and the machine learning approaches explored. The data were collected in realistic, but simulator-based scenarios. Although many prior studies support the validity of simulators for assessing behavioral and physiological changes associated with driver distraction (e.g., Eriksson, Banks, & Stanton, 2017; Mullen, Charlton, Devlin, & Bedard, 2011), several have observed differences in physiological and driver behavior measures between simulators and on road environments (e.g., Engström et al., 2005). Thus, one should be cautious about generalizing these findings beyond the task or environment explored here. Another limitation lies in the quality of data received from the sensors. Most of the data exclusions in this analysis were due to failures in the physiological sensor hardware or software. Although this in some ways strengthens the argument against a broad reliance on such sensors in distraction detection, it also limits the conclusions which can be made from these results. Finally, despite taking multiple precautions, there is some risk of model overfitting given the size of the training data and the number of parameters explored.

These limitations may be addressed through subsequent naturalistic driving studies that include a broader set of distractions, more-varied driving scenarios, and more opportunities to reliably collect accurate physiological data. Future work should also explore additional optimization techniques and more recent machine learning approaches, such as deep learning neural networks.

Application

The findings can be applied to both future technology development and empirical study design. Development of distraction mitigation systems should leverage these results to focus on driver behavior–based algorithms, in particular emphasizing lane position, speed, and steering behavior as input, as well as encouraging SVM or RF approaches and the use of diverse feature sets. Future empirical studies should consider the inclusion of lane offset standard deviation as well as quantile-based and nonlinear steering and speed metrics as dependent measures of distraction.

CONCLUSION

This study developed and tested algorithms for classifying texting, cognitive distraction, and normal driving using a mix of driver behavior and physiological input measures, a comprehensive set of feature types, and several machine learning approaches. The findings suggest that driver behavior metrics, combined with RF or SVM methods, and a diverse feature set are most promising for classifying distraction.

KEY POINTS

- Driver behavior and physiological measures may be used to create distraction classification algorithms.

- Overall, algorithms using driver behavior significantly outperform algorithms based solely on physiological measures.
- Physiological algorithms may be effective for identifying cognitive distraction but are not effective for detecting texting.
- Standard deviation of lane offset is the most sensitive feature to distraction along with speed and steering quantile measures.

ORCID iDS

Anthony D. McDonald https://orcid.org/0000-0001-7827-8828
Tyler A. Wiener https://orcid.org/0000-0003-2094-9638

REFERENCES

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, *29*, 1165–1188. doi:10.1214/aos/1013699998

Boer, E. R. (2000). Behavioral entropy as an index of workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*, 125–128. doi:10.1177/154193120004401702

Brookhuis, K. A., & de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis and Prevention*, *42*, 898–903. doi:10.1016/j.aap.2009.06.001

Caird, J. K., Johnston, K. A., Willness, C. R., Asbridge, M., & Steel, P. (2014). A meta-analysis of the effects of texting on driving. *Accident Analysis & Prevention*, *71*, 311–318. doi:10.1016/j.aap.2014.06.005

Caird, J. K., Simmons, S. M., Wiley, K., Johnston, K. A., & Horrey, W. J. (2018). Does talking on a cell phone, with a passenger, or dialing affect driving performance? An updated systematic review and meta-analysis of experimental studies. *Human Factors*, *60*, 101–133. doi:10.1177/0018720817748145

Christ, M., Kempa-Liehr, A. W., & Feindt, M. (2016). Distributed and parallel time series feature extraction for industrial big data applications. Retrieved from http://arxiv.org/abs/1610.07717

Collet, C., Guillot, A., & Petit, C. (2010). Phoning while driving I: A review of epidemiological, psychological, behavioural and physiological studies. *Ergonomics*, *53*, 589–601. doi:10.1080/00140131003672023

Cooper, J. M., Medeiros-Ward, N., & Strayer, D. L. (2013). The impact of eye movements and cognitive workload on lateral position variability in driving. *Human Factors*, *55*, 1001–1014. doi:10.1177/0018720813480177

Dietterich, T. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, *10*, 1895–1923. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9744903

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, 113, 2636–2641. doi:10.1073/pnas.1513271113

Dong, Y., Hu, Z., Uchimura, K., & Murayama, N. (2011). Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Transactions on Intelligent Transportation Systems*, *12*, 596–614. doi:10.1109/TITS.2010.2092770

Drews, F. A., Yazdani, H., Godfrey, C. N., Cooper, J. M., & Strayer, D. L. (2009). Text messaging during simulated driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51*, 762–770. doi:10.1177/0018720809353319

Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*, 97–120. doi:10.1016/j.trf.2005.04.012

Engström, J., Markkula, G., Victor, T., & Merat, N. (2017). Effects of cognitive load on driving performance: The cognitive control hypothesis. *Human Factors*, *59*, 734–764. doi:10.1177/0018720817690639

Engström, J., & Victor, T. W. (2009). Real-time distraction countermeasures. In M. Regan, J. Lee, & K. L. Young (Eds.), *Driver distraction: Theory effects and mitigation* (pp. 465–484). Boca Raton, FL: CRC Press.

Eriksson, A., Banks, V. A., & Stanton, N. A. (2017). Transition to manual: Comparing simulator with on-road control transitions. *Accident Analysis & Prevention*, *102*, 227–234. doi:10.1016/j.aap.2017.03.011

Ersal, T., Fuller, H. J. A., Tsimhoni, O., Stein, J. L., & Fathy, H. K. (2010). Model-based analysis and classification of driver distraction under secondary tasks. *IEEE Transactions on Intelligent Transportation Systems*, *11*, 692–701. doi:10.1109/TITS.2010.2049741

Fawcett, T. (2004). ROC graphs: Notes and practical considerations for researchers. *Machine Learning*, *31*, 1–38.

Feng, F., Bao, S., Sayer, J. R., Flannagan, C., Manser, M., & Wunderlich, R. (2017). Can vehicle longitudinal jerk be used to identify aggressive drivers? An examination using naturalistic driving data. *Accident Analysis & Prevention*, *104*, 125–136. doi:10.1016/j.aap.2017.04.012

Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J. G., & St John, M. (2008). The red-line of workload: Theory, research, and design. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*, 1204–1208. doi:10.1177/154193120805201811

Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the roc curve for multiple class classification problems. *Machine Learning*, *45*, 171–186. doi:10.1023/A:1010920819831

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, *6*, 156–166. doi:10.1109/TITS.2005.848368

Hicks, T. G., & Wierwille, W. W. (1979). Comparison of five mental workload assessment procedures in a moving-base driving simulator. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *21*, 129–143. doi:10.1177/001872087902100201

Hintze, J. L., & Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *The American Statistician*, *52*, 181–184. doi:10.1080/00031305.1998.10480559

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*, 196–205. doi:10.1518/001872006776412135

Jin, L., Niu, Q., Hou, H., Xian, H., Wang, Y., & Shi, D. (2012). Driver cognitive distraction detection using driving performance measures. *Discrete Dynamics in Nature and Society, 2012*, Article 432634. doi:10.1155/2012/432634

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab—An S4 package for kernel methods in R. *Journal of Statistical Software*, *11*. doi:10.18637/jss.v011.i09

Kim, S., Chun, J., & Dey, A. K. (2015). Sensors know when to interrupt you in the car. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems—CHI '15* (pp. 487–496). doi:10.1145/2702123.2702409

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *The Journal of Emergency Medicine*, *46*, 600–601. doi:10.1016/j.jemermed.2014.02.017

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, *31*, 249–268. doi:10.1115/1.1559160

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (1st ed.). New York, NY: Springer.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., . . . Hunt, T. (2017). *Package "caret": Classification and regression training*. Retrieved from https://cran.r-project.org/web/packages/caret/caret.pdf

Lee, J. D., Moeckli, J., Brown, T. L., Roberts, S. C., Schwarz, C. W., Yekhshatyan, L., . . . Davis, C. (2013). *Distraction detection and mitigation through driver feedback appendices* (Report No. DOT HS 811 547A). Washington, DC: National Highway Traffic Safety Administration.

Li, N., Jain, J. J., & Busso, C. (2013). Modeling of driver behavior in real world scenarios using multiple noninvasive sensors. *IEEE Transactions on Multimedia*, *15*, 1213–1225. doi:10.1109/TMM.2013.2241416

Liang, Y., & Lee, J. D. (2010). Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis & Prevention*, *42*, 881–890. doi:10.1016/j.aap.2009.05.001

Liang, Y., & Lee, J. D. (2014). A hybrid Bayesian network approach to detect driver cognitive distraction. *Transportation Research Part C: Emerging Technologies*, *38*, 146–155. doi:10.1016/j.trc.2013.10.004

Liang, Y., Lee, J. D., & Reyes, M. L. (2007). Nonintrusive detection of driver cognitive distraction in real time using Bayesian networks. *Transportation Research Record: Journal of the Transportation Research Board*, *2018*, 1–8. doi:10.3141/2018-01

Liang, Y., Reyes, M. L., & Lee, J. D. (2007). Real-time detection of driver cognitive distraction using support vector machines. *IEEE Transactions on Intelligent Transportation Systems*, *8*, 340–350. doi:10.1109/TITS.2007.895298

Liaw, A., & Weiner, M. (2002). Classification and regression by random forest. *R News*, 18–22. Retrieved from http://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf

Liu, T., Yang, Y., Huang, G., Yeo, Y. K., & Lin, Z. (2016). Driver distraction detection using semi-supervised machine learning. *IEEE Transactions on Intelligent Transportation Systems*, *17*, 1108–1120. doi:10.1109/TITS.2015.2496157

Macdonald, W. A., & Hoffmann, E. R. (1980). Review of relationships between steering wheel reversal rate and driving task demand. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *22*, 733–739. doi:10.1177/001872088002200609

Majka, M. (2018). *naivebayes: High performance implementation of the naive bayes algorithm*. Retrieved from https://cran.r-project.org/package=naivebayes

Masood, S., Rai, A., Aggarwal, A., Doja, M. N., & Ahmad, M. (2018). Detecting distraction of drivers using convolutional neural network. *Pattern Recognition Letters*. Advance online publication. doi:10.1016/j.patrec.2017.12.023

Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board*, *2138*, 6–12. doi:10.3141/2138-02

Mehler, B., Reimer, B., & Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*, 396–412. doi:10.1177/0018720812442086

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C., & Lin, C.-C. (2017). *R Package e1071 Version 1.6-8. Gpl-2*. Retrieved from https://cran.r-project.org/web/packages/e1071/e1071.pdf

Miyaji, M., Kawanaka, H., & Oguri, K. (2009). Driver's cognitive distraction detection using physiological features by the AdaBoost. In *IEEE conference on intelligent transportation systems (ITSC)* (pp. 90–95). doi:10.1109/ITSC.2009.5309881

Mullen, N., Charlton, J., Devlin, A., & Bedard, M. (2011). Simulator validity: Behaviors observed on the simulator and on the road. In D. L. Fisher, M. Rizzo, J. K. Caird, & J. D. Lee (Eds.), *Handbook of driving simulation for engineering, medicine, and psychology* (pp. 1–18). Boca Raton, FL: CRC Press.

National Center for Statistics and Analysis. (2017, March). *Distracted driving 2015*. (Traffic Safety Facts Research Note. Report No. DOT HS 812 381). Washington, DC: National Highway Traffic Safety Administration.

Pavlidis, I., Dcosta, M., Taamneh, S., Manser, M., Ferris, T., Wunderlich, R., & Tsiamyrtzis, P. (2016). Dissecting driver behaviors under cognitive, emotional, sensorimotor, and mixed stressors. *Scientific Reports*, *6*, 1–12. doi:10.1038/srep25651

Ragab, A., Craye, C., Kamel, M. S., & Karray, F. (2014, October). *A visual-based driver distraction recognition and detection using random forest*. Paper presented at the International Conference Image Analysis and Recognition, Vilamoura, Portugal (pp. 256–265).

R core team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org

Reimer, B., Mehler, B., Coughlin, J. F., Roy, N., & Dusek, J. A. (2011). The impact of a naturalistic hands-free cellular phone task on heart rate and simulated driving performance in two age groups. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*, 13–25. doi:10.1016/j.trf.2010.09.002

Rodriguez-Paras, C., Susindar, S., Lee, S., & Ferris, T. K. (2017). Age effects on drivers' physiological response to workload. *Proceedings of the Human Factors and Ergonomics Society, 61*, 1886. doi:10.1177/1541931213601951

Rodriguez-Paras, C., Yang, S., Tippey, K., & Ferris, T. K. (2015). Physiological indicators of the cognitive redline. *Proceedings of the Human Factors and Ergonomics Society*, *59*, 637–641. doi:10.1177/1541931215591139

Sathyanarayana, A., Nageswaren, S., Ghasemzadeh, H., Jafari, R., & Hansen, J. H. L. (2008). Body sensor networks for driver distraction identification. In *2008 IEEE international conference on vehicular electronics and safety* (pp. 120–125). New York, NY: IEEE. doi:10.1109/ICVES.2008.4640876

Schliep, K., & Hechenbichler, K. (2016). *kknn: Weighted k-nearest neighbors*. Retrieved from https://cran.r-project.org/package=kknn

Schreiber, T., & Schmitz, A. (1997). Discrimination power of measures for nonlinearity in a time series. *Physical Review E*, *55*, 5443–5447. doi:10.1103/PhysRevE.55.5443

Schwarz, C., Brown, T., Lee, J., Gaspar, J., & Kang, J. (2016). The detection of visual distraction using vehicle and driver-based sensors. In *SAE 2016 world congress and exhibition*, Detroit, MI. doi:10.4271/2016-01-0114

Smith, M. R. H., Witt, G. J., Bakowski, D. L., Leblanc, D., & Lee, J. D. (2009). Adapting collision warnings to real-time estimates of driver distraction. In M. Regan, J. D. Lee, & K. L. Young (Eds.), *Driver distraction: Theory effects and mitigation* (pp. 501–518). Boca Raton, FL: CRC Press.

Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., & Mehler, B. (2014). Classifying driver workload using physiological and driving performance data. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems—CHI '14* (pp. 4057–4066). New York, NY: ACM Press. doi:10.1145/2556288.2557068

Son, J., & Park, M. (2016, May). *Real-time detection and classification of driver distraction using lateral control performance*. Paper presented at the ACCSE 2016: The First International Conference on Advances in Computation, Communications and Services, Valencia, Spain.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied*, *9*, 23–32. doi:10.1037/1076-898X.9.1.23

Taamneh, S., Tsiamyrtzis, P., Dcosta, M., Buddharaju, P., Khatri, A., Manser, M., . . . Pavlidis, I. (2017). A multimodal dataset for various forms of distracted driving. *Scientific Data*, *4*, Article 170110. doi:10.1038/sdata.2017.110

Therneau, T., Atkinson, B., & Ripley, B. (2017). *rpart: Recursive partitioning and regression trees*. Retrieved from https://cran.r-project.org/package=rpart

Tippey, K. G., Sivaraj, E., & Ferris, T. K. (2017). Driving while interacting with Google glass: investigating the combined effect of head-up display and hands-free input on driving safety and multitask performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*, 671–688. doi:10.1177/0018720817691406

Torkkola, K., Massey, N., & Wood, C. (2004). Detecting driver inattention in the absence of driver monitoring sensors. In *Proceedings of the 2004 International Conference on Machine Learning and Applications* (pp. 220–226). New York, NY: IEEE. doi:10.1109/ICMLA.2004.1383517

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York, NY: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (4th ed.). New York, NY: Routledge.

Zhang, Y., Owechko, Y., & Zhang, J. (2004). Driver cognitive workload estimation: a data-driven perspective. In *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems* (IEEE Cat. No. 04TH8749, pp. 642–647). New York, NY: IEEE. doi:10.1109/ITSC.2004.1398976

Anthony D. McDonald is an assistant professor of industrial and systems engineering at Texas A&M University and directs the Human Factors and Machine Learning Laboratory. He received his PhD in industrial engineering from the University of Wisconsin–Madison in 2014.

Thomas K. Ferris is an associate professor of industrial and systems engineering at Texas A&M University, where he is the director of the Human Factors and Cognitive Systems Laboratory. He received his PhD in industrial and operations engineering from the University of Michigan in 2010.

Tyler A. Wiener is a supply chain engineer at Walmart Incorporated. He earned his BS in industrial engineering from Texas A&M University in 2018. The present paper was authored while he was an undergraduate researcher in the Human Factors and Machine Learning Laboratory at Texas A&M University.

Date received: July 23, 2018
Date accepted: May 15, 2019