Accident Analysis and Prevention 113 (2018) 25–37

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

A contextual and temporal algorithm for driver drowsiness detection

Anthony D. McDonald[a,*], John D. Lee[b], Chris Schwarz[c], Timothy L. Brown[c]

[a] Texas A&M University, Department of Industrial and Systems Engineering, 101 Bizzell Street, College Station, TX 77845, USA
[b] University of Wisconsin-Madison, Department of Industrial and Systems Engineering, 1513 University Avenue, Madison, WI 53706, USA
[c] National Advanced Driving Simulator, The University of Iowa, 2401Oakdale Blvd, Iowa City, IA 52242, USA

A R T I C L E I N F O

Keywords:
Drowsiness
Detection
Dynamic Bayesian Network
Random forest
Driver safety

A B S T R A C T

This study designs and evaluates a contextual and temporal algorithm for detecting drowsiness-related lane. The algorithm uses steering angle, pedal input, vehicle speed and acceleration as input. Speed and acceleration are used to develop a real-time measure of driving context. These measures are integrated with a Dynamic Bayesian Network that considers the time dependencies in transitions between drowsiness and awake states. The Dynamic Bayesian Network algorithm is validated with data collected from 72 participants driving the National Advanced Driving Simulator. The algorithm has a significantly lower false positive rate than PERCLOS—the current gold standard—and baseline, non-contextual, algorithms under design parameters that prioritize drowsiness detection. Under these parameters, the algorithm reduces false positive rate in highway and rural environments, which are typically problematic for vehicle-based detection algorithms. This algorithm is a promising new approach to driver impairment detection and suggests contextual factors should be considered in subsequent algorithm development processes. It may be combined with comprehensive mitigation methods to improve driving safety.

1. Introduction

The National Highway Traffic Safety Administration (NHTSA, 2011) reported drowsiness contributes to approximately 83,000 crashes, 37,000 injuries, and 900 deaths each year—accounting for approximately 3% of all traffic-related fatalities. The 100-Car naturalistic driving study found that drowsy driving contributed to 22%–24% of the crashes and near-crashes observed (Klauer et al., 2006). Crash survey data illustrate that this problem is not unique to American drivers—-drowsiness contributes to as many as 7% of crashes in the United Kingdom and 3.9% of crashes in Norway (Maycock, 1997; Sagberg, 1999). The variance in these estimates reflects the difficulty associated with identifying drowsiness-related crashes. This difficulty is driven by the fact that drowsiness leaves no physical trace and is a subjective experience. This lack of physical evidence suggests that the crash statistics and surveys may underestimate the true problem of drowsy driving.

The majority of drowsiness-related crashes, nearly 80%, can be classified as single car run off road crashes, where the driver stops controlling their vehicle and eventually departs their lane and the roadway (Pack et al., 1995). Reducing these crashes requires a multifaceted approach including schedule restrictions for professional drivers (Gander et al., 2011), increased education for drivers (Fletcher et al., 2005), laws against drowsy driving (Geist et al., 2002), driver feedback (Aidman et al., 2015), and detection and mitigation technology (Balkin et al., 2011). The role of detection and mitigation technology in this approach is to provide an intervention immediately prior to a crash that prevents the crash from occurring or reduces its severity. One specific goal of detection and mitigation technology is to detect and prevent single car run off road crashes caused by drowsiness. The scope of this goal includes both cases of prolonged and intermittent drowsiness.

Detection and mitigation technology consists of collecting data from the driver, vehicle, or environment, applying a classification algorithm to these data, and presenting the result of the classification algorithm to the driver (Balkin et al., 2011). A substantial amount of research has been dedicated to optimizing the data collection and classification algorithm application (Liu et al., 2009). The goals of this research typically center on introducing novel input measures (Dinges and Grace, 1998; Lal et al., 2003), evaluating the use of machine learning approaches that have been successful in other domains (Patel et al., 2011; Yang et al., 2010; Yeo et al., 2009), or introducing novel pre-processing steps to improve classification (Kutila et al., 2007; Sayed and Eskandarian, 2001). These three directions of research can be condensed into a discussion of algorithm input and machine learning methods.

* Corresponding author. Present address: 4075 Emerging Technologies Building, 101 Bizzell Street, College Station, Texas, 77845, USA.
E-mail address: mcdonald@tamu.edu (A.D. McDonald).

https://doi.org/10.1016/j.aap.2018.01.005
Received 20 September 2016; Received in revised form 5 January 2018; Accepted 6 January 2018
Available online 02 February 2018
0001-4575/ Published by Elsevier Ltd.

1.1. Drowsiness detection algorithm input

Drowsiness detection algorithm input sources can be differentiated by the raw measurement and the processing steps taken to convert measurements into features. Measures explored in the literature include: heart rate (Furman et al., 2008), brain activity (Dinges et al., 1998; Lal et al., 2003), eye closure and tracking (Dinges et al., 1998; Ji et al., 2004; Wierwille et al., 1994b), lane position (Hanowski et al., 2008a), and steering-wheel angle (Eskandarian and Mortazavi, 2007; Krajewski, Golz, et al., 2009; Krajewski and Sommer, 2009; Sayed and Eskandarian, 2001). Although most previous algorithms focus on a one type of measure, several employ a combination of measures (Forsman et al., 2013; Hanowski et al., 2008b; Ji et al., 2004; Tijerina et al., 1999; Zilberg et al., 2007). The most commonly applied and theoretically rigorous measures are electroencephalogram (EEG), percent eye-closure over a fixed time window (PERCLOS), and steering-wheel angle (Balkin et al., 2011). EEG is advantageous because spectral patterns in the signal have a well-established link to the transition between wakefulness and sleep (Lal and Craig, 2001). EEG is limited by the amount of pre-processing required prior to classification, vulnerability to artifacts, and the feasibility of collecting EEG from drivers in real situations. PERCLOS, developed by Wierwille et al. (1994a), is the gold standard measure for drowsiness detection. PERCLOS predicts drowsiness based on the percentage of time an individual's eyes are more than 80% closed over a 2-min period. Dinges et al. (1998) demonstrated that the PERCLOS algorithm had over 90% accuracy in detecting degraded performance during a vigilance task, which was more reliable across drivers than EEG, blinks, and head position in the study. PERCLOS has been incorporated into aftermarket devices such as the Co-pilot (Grace and Stewart, 2001) and has been used as a ground truth measure of drowsiness (Tijerina et al., 1999; Wierwille et al., 1994b). Despite its wide acceptance PERCLOS has several practical limitations. PERCLOS for real-time detection is limited because current camera technology required for its measurement is expensive, has not been extensively validated, and may be unreliable when the driver wears sunglasses or under weather conditions that produce high amounts of glare (Balkin et al., 2011). Despite these limitations, the substantial evidence showing the utility of PERCLOS suggests that it might be useful for benchmarking new algorithms.

The limits of PERCLOS and EEG have led researchers to examine steering-wheel angle, or the deflection of the top of the wheel from the zero point. Steering-wheel angle is similar to EEG data in that it requires significant pre-processing and transformation before it becomes a viable input measure. Sayed and Eskandarian (2001) introduced a steering-wheel angle based algorithm that filtered raw steering angle measure to remove road curvature events, and then discretized into bins of steering patterns. The algorithm classified drivers labeled as sleep deprived or non-sleep deprived with nearly 90% accuracy. Similarly, Krajewski et al. (2009) developed an algorithm that processed raw steering-wheel angle data into features characterized the signal in the time and frequency domains. The algorithm also included features representing the non-linear aspects of steering-wheel angle patterns. The algorithm achieved 86% accuracy in identifying self-reported sleepiness. McDonald et al. (2013a) presented an approach that used raw steering-wheel angle data, however the machine learning technique applied internally filtered the data. The algorithm performed comparably to PERCLOS in detecting drowsy-related lane departures. Steering-wheel data is limited in two primary facets. First, it is highly sensitive to differences in driving activities, such as curve negotiation, and thus detection could be confounded with differences in the driving context (Balkin et al., 2011; Hartley et al., 2000). Second, patterns in steering wheel angle that are indicative of drowsiness, namely a lack of steering input, are often also associated with other impairments such as distraction. To overcome these limitations, steering-wheel angle based algorithms accommodate road artifacts and either carefully consider the ground truth definition of drowsiness or be trained to differentiate multiple types of impairment. Algorithms that focus solely on drowsiness detection using steering wheel angle must have a ground truth definition that clearly differentiates between distraction and drowsiness.

1.2. Machine learning methods in drowsiness detection algorithms

Machine learning methods can be characterized by their training procedure, prediction procedure, and their optimization parameters (Kotsiantis et al., 2007). The drowsiness detection literature has explored a variety of methods including: Decision Trees (Krajewski and Sommer, 2009; McDonald et al., 2013b), Neural Networks (Eskandarian and Mortazavi, 2007; Garcés Correa et al., 2014; Patel et al., 2011; Sandberg et al., 2011; Sayed and Eskandarian, 2001; Vuckovic et al., 2002; Wilson and Bracewell, 2000), Support Vector Machines (Awais et al., 2017; Hu and Zheng, 2009; Jo et al., 2014; Kutila et al., 2007; Lee et al., 2015; Sun et al., 2017; Zhao et al., 2012), Logistic Regression (Murata, 2016), Random Forests (McDonald et al., 2013b; Wang et al., 2016), Bayesian Networks (Ji et al., 2004; Yang et al., 2009), and Dynamic Bayesian Networks (Ji et al., 2006; Yang et al., 2010; Yang et al., 2009). Of these approaches Dynamic Bayesian Networks (DBN) are the most promising for future work because they explicitly model the time-dependent nature of driver drowsiness and allow the inclusion of contextual factors that influence drowsy driving, such as prior sleep behavior and road type. DBN models consist of graph structures—nodes connected by directed edges—that mimic the dependencies in the underlying problem, and an associated group of probabilities that model the likelihood of model state transitions. The dynamic components of the model specify dependencies across time (Murphy, 2002). More specifically, DBN algorithms can encode facts about drowsiness such as drivers that are drowsy are likely to stay drowsy and that drivers that are awake tend to stay awake. Specification of a DBN requires indicating the probabilities or probability distributions that characterize the relationships in the model.

Several studies have explored the utility of DBN for drowsiness detection. Ji et al. (2006) developed an algorithm that combines contextual, facial, ocular, and head-position input to predict drowsiness as defined by reaction times during a non-driving vigilance task. The contextual information in the algorithm consisted of circadian rhythm, sleep quality, the presence of sleep disorders, and information about work environments. The probability distributions for these contextual factors were inferred based on domain knowledge. Yang et al. (2010) extended the work by adding heart rate, EEG, and eye measures as input to the previous algorithm. While these studies demonstrate the potential effectiveness of the DBN framework for detecting drowsiness, they carry many of the same limitations associated with other EEG and eye-closure based algorithm and they do not consider relevant contextual aspects in drowsy driving, such as the type of road and driving maneuvers (e.g. lane changes). The studies discussed in this review are summarized in Table 1. A more thorough review can be found in Lenné and Jacobs (2016).

1.3. A temporal and contextual algorithm for drowsiness detection

The role of the type of road in drowsiness related crash risk is well established (I. D. Brown, 1994). Crashes attributed to drowsiness are significantly more common on rural straight roads that do not contain sufficient stimuli to keep the driver awake. Furthermore many studies and models of driver behavior illustrate that drivers alter their driving behavior relative to the driving context (McRuer et al., 1977; Michon, 1986; Salvucci, 2006; Weir and McRuer, 1970; Wilde, 1988). The significance of context in both unimpaired and drowsy driving behavior suggests there is a gap in the literature for drowsiness detection algorithms that include on-road contextual factors as an input. These factors may include both the type of road (residential street, highway, urban arterial) and the immediate environment around the driver (other

Table 1
Summary of current drowsiness detection algorithms arranged according to the type of study, algorithm input, and machine learning method. Rows labeled with "continuous measure analysis" indicate that the study focused on correlative analysis of continuous variables rather than discrete predictions.

Test environment	Study authors	Algorithm input	Machine learning method
Driving Simulation	Furman et al. (2008)	Heart Rate Variability	Decision stump or threshold
	Lal et al. (2003)	EEG	Continuous measure analysis
	Eskandarian and Mortazavi (2007)	Steering Wheel Input	Neural network
	Krajewski et al. (2009), Krajewski and Sommer (2009)	Steering Wheel Input	Ensemble classification
	Sayed and Eskandarian (2001)	Steering Wheel Input	Neural network
	Forsman et al. (2013)	Steering Wheel and Lane position metrics	Continuous measure analysis
	Wierwille et al. (1994b)	PERCLOS	Continuous measure analysis
	Zilberg et al. (2007)	Various physiological metrics	Continuous measure analysis
	McDonald et al. (2013b)	Steering Wheel Input	Random forest
	Patel et al. (2011)	Heart Rate Variability	Neural network
	, Sandberg et al. (2011)	Various driving behavioral metrics	Neural network
	Yang et al. (2009)	Task performance	Bayesian network and DBN
	Yang et al. (2010)	Environmental factors, Eye closure, head movement, facial expressions, EEG, Heart rate measures	DBN
	Lee et al. (2015)	Smart watch based physiological and steering measures	Support Vector Machine
	Wang et al. (2006)	Longitudinal and Lateral acceleration of the vehicle, steering	Random forest
	Van Loon et al. (2015)	Lateral position and steering	Continuous measure analysis
	Murata (2016)	EEG, Heart Rate, Eye Closure, Foot and back pressure, lane tracking error, neck bending angle	Logistic regression
	Wang et al. (2017)	Driver factors and Environmental conditions	Accelerated failure time model
	Figure 2	EEG, Heart Rate Measures	Support Vector Machine
Naturalistic driving	Jo et al. (2014)	Eye closure based metrics	Support Vector Machine
	Sun et al. (2017)	Environmental factors, Eye closure metrics, Various vehicle based measures	Multi-class SVM and Decision-level fusion
	Tijerina et al. (1999)	PERCLOS, various vehicle based metrics	Decision stump or threshold
Controlled naturalistic driving study	Garcés Correa et al. (2014)	EEG	Neural Network
Non-driving lab study	Dinges and Grace (1998), Dinges et al. (1998)	PERCLOS	Decision stump or threshold
	Ji et al. (2004)	Environmental factors, Eye closure, head movement, facial expressions	Bayesian network
	Ji et al. (2006)	Environmental factors, Eye closure, head movement, facial expressions	DBN

vehicles, road condition, lane width and markings).

Like driving context, time plays a critical role in drowsiness related crashes. Temporal factors such as the time of day, circadian nadirs, and continuous driving time are often seen as contributing factors to drowsiness related crashes (I. D. Brown, 1994; Pack et al., 1995). Temporal dependencies also exist at smaller scales; from 1 min to the next alert drivers will likely stay alert and drowsy drivers will stay drowsy. Temporal modeling approaches, such as DBN, can be used incorporate these dependencies into detection algorithms. Prior work in this area suggests that DBNs are a promising direction (Ji et al., 2006; Yang et al., 2010), however there is a gap in the current literature for vehicle-based DBN algorithms.

This study developed a temporal and contextual algorithm for drowsiness-related lane departure detection. The contextual components of this algorithm improve on current algorithm performance because they differentiate contexts where a drowsiness-related crash is not likely such as in an urban or suburban setting. They may also contribute to addressing the known limitation of driver-behavioral based algorithms where the algorithm confuses drowsy driving with driving on a straight road (Balkin et al., 2011; McDonald et al., 2013b). The remaining sections of this paper discuss the data used to develop this algorithm, present a novel method of road-context input generation, illustrate the model selection process, and evaluate the algorithm relative to simpler algorithms and PERCLOS.

2. Material and methods

This section reviews the dataset, ground truth definition of drowsiness, and feature generation process used in this study. This dataset was collected at the National Advanced Driving Simulator (NADS) located on the campus of the University of Iowa (NADS, 2010), ground truth drowsiness is defined by the presence of one or more drowsiness-related lane departures in a 60s window, and the feature generation process consisted of data processing with a random forest algorithm and Symbolic Aggregate Approximation (SAX; Lin et al., 2007).

2.1. Simulator data

The NADS simulator is a high fidelity driving simulation environment consisting of a 7.3 m dome containing a full Chevy Malibu sedan surrounded by 360 degrees of screens. The dome is located on a full motion base that provides 400 m of lateral and longitudinal travel and 330 degrees of rotation in both directions. Data were collected from seventy-two participants as they completed three drives of approximately 30 min each. The drivers were healthy men and women from one of three age groups: 21–34, 38–51, 55–68. All participants possessed a valid United States driver's license, were confirmed healthy during a physical and medical history examination, and did not tend strongly toward wakefulness in the evening hours. Participants were compensated $250 for study participation or a prorated amount if they did not complete the study. In addition to good health, participants had to have driven a minimum of 10,000 miles per year for the past two years, live within a 30-min drive of the NADS office, have sleep patterns in which they slept and woke at approximately the same time every day, and were asked to abstain from caffeine after 1200 on the day of the overnight visit. Prior to each drive participants were screened for drug use and alcohol consumption, and were removed from the study if they tested positive. Participants' sleep was monitored with an actigraph and participants who slept less than six hours on the night before the drive were excluded from the study.

2.1.1. Data collection

Data were collected continuously throughout each drive from the NADS simulator and a dashboard-mounted eye-tracker (Face Lab™ 5.0, Seeing Machines, Canberra, Australia). The simulator collects a record of vehicle state (e.g. lane position and speed) and driver inputs (e.g. steering wheel position and accelerator pedal position) originally sampled at 240 Hz. The eye-tracker records a variety of eye measures including gaze and closure, sampled at 60 Hz. In addition to these sensors, EEG data and driver postural data were collected. To balance the data and facilitate data sharing the driving data were down-sampled to 60 Hz prior to analysis.

2.1.2. Study procedure

The study period consisted of three separate visits to the simulator. The first visit was a screening visit and the second and third visits consisted of either a single daytime drive or two drives during the late evening (Early Night condition) and early morning (Late Night condition) respectively. The order of these days was counterbalanced across participants and visits were separated by a period of at least three days. Prior to the start of each drive, participant's drowsiness was evaluated with the Stanford Sleepiness Scale (SSS; Hoddes et al., 1973) and a modified Psychomotor Vigilance Test (PVT; Wilkinson and Houghton, 1982). The SSS is a subjective 1–7 rating scale of drowsiness and the PVT is a visual stimulus, manual response task that measures reaction time. After each driver completed the SSS and PVT a second time and also rated their drowsiness during various points in the drive using a Retrospective Sleepiness Rating. In the night-time session participants were kept awake between the drives via study personnel intervention and various activities including television and books.

The daytime drive began between 0900 and 1200, the early night drive began between 2200 and 0100, and the late night drive began between 0200 and 0500. The late night drive occurred after a minimum of 18 h of continuous wakefulness. Each drive consisted of three connected segments representing urban, highway, and rural environments. The drive began with an urban segment on a two-lane roadway with posted speed limits between 25 and 45 mph. This segment contained several controlled and uncontrolled intersections along with a series of potential hazards that consisted of other vehicles, motor-bikes, and pedestrians entering the roadway. Following the urban segment drivers entered a four-lane divided expressway with a posted speed limit of 70 mph. During this segment, drivers followed a lead vehicle and then overtook a series of slower vehicles. After exiting the highway, participants drove on a rural, undivided, two-lane road that culminated in a 300 s period of continuous straight driving. The goal of this drive was to simulate a drive home from an urban parking spot to a rural home location. The drive was designed to represent a typical nightlight drive, with an emphasis on road situations associated with drowsiness-related crashes, specifically the highway and rural driving components. The drive was approximately 35 min long and contained realistic driving surroundings throughout. To eliminate a potential learning effect three slightly different driving scenarios were used one for each drive. Each scenario presented the same simulator events but in a different order. The contents of each drive were categorized into events associated with a specific event name, shown in Table 2. More details on the study can be found in T.L. Brown et al. (2014).

2.1.3. Training and Testing Data

A key component of any machine learning process is the partition of training and testing data. Training data are used to develop the algorithm and make decisions regarding feature selection, model structure, and model parameter selection. Test data are held apart for these decision and thus provide an objective measure of the algorithm's detection performance. In this study, training and test data were partitioned through random sampling. The test data consisted of 10% of the original data. Data were partitioned at the participant level to avoid any bias associated with individual driving norms. In total, 7 participants were included in the test dataset and 65 were included in the training data. The goal of the random sample was to achieve as much balance across demographics and experiment conditions as possible.

The final testing dataset contained a representative from every experimental condition. Table 3 shows a demographic summary of the final testing dataset. This dataset contains 4 female drivers and 3 male drivers, ranging in age from 22 to 57. Five of the drivers began data collection with the early and late night drives and two drivers began with the daytime drive. The testing data were withheld from all feature and parameter selection processes.

2.2. Ground truth definition

The choice of the ground truth definition of drowsiness is critical to the success of a drowsiness mitigation system (Balkin et al., 2011). Detection and mitigation systems that include algorithms trained to detect broad measures of ground truth such as hours of wakefulness may be prone to providing warnings that are viewed by drivers as either non-informative or false positives. These "misclassifications" may lead to distrust and a reduced use of the system (Ghazizadeh et al., 2012; Lee and See, 2004). In order to overcome such issues, this study defines ground truth drowsiness as drowsy-related lane departures. These drowsy-related lane departures are different from a routine lane departure, which could be caused by inattention, correction failures, or other impairments, in that they are preceded by a 1-min period of driving in which the driver displayed visible drowsiness indicators such as eye closures, yawning, head nodding, and slumping posture. Drowsy-related lane departures were selected as a ground truth definition of drowsy driving because they are generalizable to all drivers, are easy to communicate to drivers, are clearly related to the consequences of drowsiness, and might motivate drivers to pursue action.

Drowsy-related lane departures were identified through video analyses of all lane departures observed in the study. During this analysis erroneous lane departures caused by simulator anomalies and visual-manual distraction were reclassified as non-drowsy periods. Each departure was manually coded using the Observer Rating of Drowsiness (ORD) scale. The inter-rater reliabilities—measured by inter-class correlation coefficient (ICC)—of the raters were 0.69 and 0.72. The ORD scale is a continuous rating between 0 and 100, based on the 60 s preceding each lane departure (Wierwille et al., 1994b). In this case the ORD measures were separated into five bins: not drowsy ($ORD < 12$), slightly drowsy ($12 \le ORD < 37$), moderately drowsy ($37 \le ORD < 62$), very drowsy ($62 \le ORD < 90$), and extremely drowsy ($ORD \ge 90$). Departures classified as moderately, very, or extremely drowsy were labeled "drowsy."

One limitation of drowsy-related lane departures is that they occur on the order of a second, whereas the patterns of driver behavior associated with them often occur on the order of a minute. This study accounted for this issue by dividing data into time windows and labeling each window as drowsy or awake based on the presence of a drowsiness-related lane departure within the window. Windows where the driver recognized the lane departure and corrected for it were categorized as awake instances. This process is illustrated in Fig. 1, which shows the steering and brake pedal input surrounding the drowsy-related lane departures observed in the training data. Windows containing no corrective action (the plots on the left hand side of each chart) were classified as drowsy. Windows where the participant performed a corrective action (the right side of each chart) were classified as awake.

2.3. Algorithm evaluation

This study explored two approaches to algorithm evaluation, each supported with a separate series of statistical evaluations. The first approach of algorithm evaluation is based on the area under the Receiver Operating Characteristic (ROC) curve (AUC). The ROC curve is

Table 2
Event descriptions for the simulator study.

Road Environment	Event Short Name	Description
Urban	Pullout	Pull out of parallel parking space
	Urban Arterial	Driving on a narrow urban road with parked cars on both sides
	Green	Navigating through a green light at a controlled intersection
	Yellow Dilemma	A yellow light dilemma at a controlled intersection
	Left	A left turn at a controlled intersection
	Urban Curves	Navigating a series of curves on an urban two-lane road with cars parked on both sides
	On Ramp	Navigating a highway entrance ramp
Highway	Merge On	Merging on to the highway
	Interstate	Driving behind a slow moving vehicle on the highway
	Merging Traffic	Driver approaches an interchange with a vehicle merging about 500 feet ahead of the on-ramp
	Interstate Curves	A series of three curves the driver must negotiate on the interstate with light traffic
	Exit Ramp	Navigating a highway exit ramp
	Turn Off Ramp	Right turn from the off-ramp onto a rural two-lane undivided road
Rural	Lighted	Driving on a lighted two-lane rural road with a speed limit of 55 mph
	Transition To Dark	Transition to a segment of the rural road that is unlighted
	Dark	Driving on a rural, two-lane, unlighted 55 mph road with faded lane lines involving some curves
	Transition to Gravel	Turn slightly to the right onto a gravel road
	Gravel	Driving on the gravel road
	Driveway	Pulling in to a gravel driveway
	Gravel Rural Ext.	Navigating an unlighted gravel rural road that contains a series of curves
	Gravel Transition To Straight	Transitioning to a straight segment of gravel road.
	Paved Transition To Rural Straight	Driving on a transition from a gravel road to a paved rural road
	Rural Straight	Navigating an unlighted paved rural road for 10 min
	Dark No Hairpin	Driving in the dark on a straight rural road
	Dark Hairpin	Driving in the dark on a hairpin turn on a rural road

Table 3
Demographics and drowsy-related lane departure frequency for the drivers in the test dataset.

Participant	Age	Gender	Order
1	22	F	Night First
2	55	F	Day First
3	57	M	Day First
4	57	M	Night First
5	56	F	Night First
6	47	M	Night First
7	50	F	Night First

a plot of the true positive rate by the false positive rate of an algorithm for a range of thresholds. The threshold represents the minimum evidence required for the algorithm to predict that a driver is drowsy. For example, in a Dynamic Bayesian Network algorithm the threshold is based on probability of drowsiness output from the algorithm. The AUC is a robust metric of algorithm evaluation because it is insensitive to the underlying class distribution (Fawcett, 2004). AUC can be calculated for an entire ROC curve or for partial ROC curves over a range of false positive rates. Differences AUC values can be statistically evaluated with a bootstrapped significance tests (Robin et al., 2011). The AUC provides a general comparison between algorithms independent of decision threshold.

The limitation of the AUC is that in practice algorithms require the use of only a single threshold. This threshold represents a design decision that balances the importance of true and false positives. Higher thresholds will detect fewer instances of drowsiness but also show fewer false positives whereas lower thresholds will detect more instances of true drowsiness but also have more false positives. Thus it is beneficial to supplement AUC comparisons with evaluations aligned with common design preferences. This type of analysis produces specific true and false positive rates. These values can be statistically evaluated using a McNemar's test (Dietterich, 1998), or a Fisher's exact test of count data for smaller samples ($n < 25$). The goal of these tests is to evaluate whether or not there are practical settings of various algorithms at

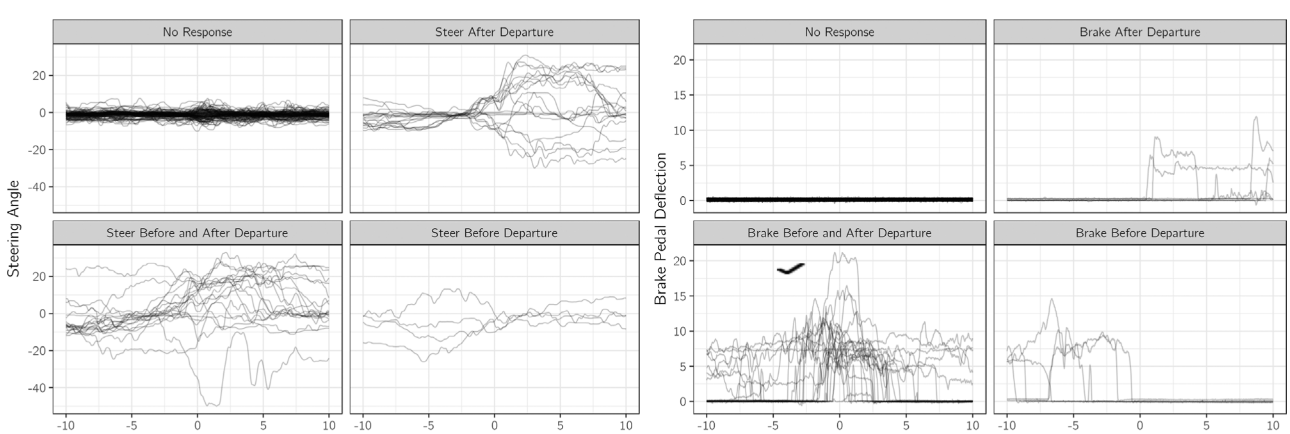

Fig. 1. Steering and brake input in the 10 s surrounding a drowsy-related lane departure faceted by the corrective action. A braking action is defined as a pedal input of greater than 1 degree and steering action is defined as a steering wheel angle greater than 5 degrees. Note that the plots are centered on the lane departure and the lack of corrective action is significantly more frequent than any type of corrective action.

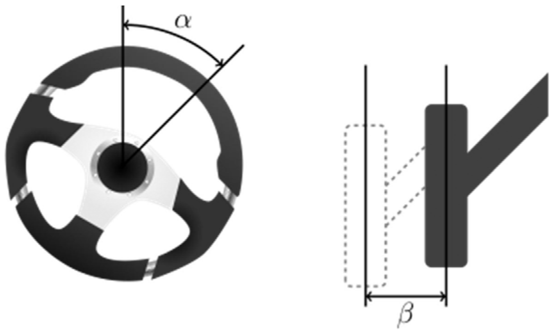

Fig. 2. Illustration of steering wheel angle (α) and pedal deflection (β).

which one algorithm performs significantly better than another.

2.4. Feature generation

The modeling approach discussed in this study used two types of features: driver behavioral measures and road context measures. Both of these required pre-processing to be converted from their raw form to algorithm input. While there are many ways to approach this problem, this study focused on generating driver behavioral features through pre-processing with a random forest algorithm and developed driving context via Symbolic Aggregate Approximation (SAX). These methods are advantageous because they are not computationally intensive and there is some evidence to suggest that they adequately capture relevant aspects of driver input and context respectively (McDonald et al., 2017; McDonald et al., 2013a; McDonald et al., 2013a).

2.4.1. Driver behavior feature generation

The driver behavior measures used in this algorithm consist of steering wheel angle, accelerator pedal input, and brake pedal input. Steering-wheel angle, shown in the left side of Fig. 2, refers to the angle between the current location of the vertical center of the steering wheel and the resting position of the wheel. Similarly pedal input (right side of Fig. 2) is defined by the distance between the pedal's current position relative to the neutral position.

Post-processing of steering and pedal data can be accomplished through data reduction techniques such as Fourier Transforms, Distributional statistics (mean, median, mode, min, max, skew, kurtosis), and through applying machine learning methods. The drowsiness detection algorithm literature typically views machine learning methods as the recipients of features rather than the feature generation technique; however, any machine learning method that can produce a continuous output (e.g. votes from a random forest) can also be used as a feature in a broader algorithm. Using the output of one algorithm as features for another produces a hierarchical structure where the machine-learning feature generation method filters likely instances of drowsiness and the broader algorithm aggregates across machine learning methods or, as in this study, biases them based on a model of the environment. Further examples of this approach can be found in Krajewski et al. (2009) and Sun et al. (2017).

This study explored all three methods and assessed them relative to their information gain on the training data. Information gain feature filtering is a common practice in the machine learning literature used to identify input measures that have high predictive value and remove input measures that have either low predictive value or are underrepresented in the data (Guyon and Elisseeff, 2003). The comparison included features generated via Fourier Transform, distributional statistics, a neural network, a Support Vector Machine, a k-Nearest Neighbor model, a Naïve Bayes classifier, a Decision Tree, and a random forest. The features were created from 60 s steering and pedal input sampled at 60 Hz. The machine learning methods used distributional statistics as their input. The results of the information gain analysis showed that the features generated from the random forest method were the most powerful for detecting drowsiness, i.e. they had the highest information gain for both steering and pedal input. This result aligns with a similar exploration in McDonald et al. (2013a) that showed the random forest algorithm outperformed other machine learning approaches in using steering angle data to detect drowsiness related-lane departures. Based on the results of this comparison, the random forest features were retained and the remaining feature types were removed from further analyses.

2.4.2. Contextual feature generation with symbolic aggregate approximation

The concept of road context is intuitive in the sense that one could narrate it to an observer. For example, "I am currently driving in the right lane on a highway in clear weather. There is a vehicle about two car lengths behind me and one in the left lane passing me." Despite its intuitiveness it is difficult to capture such data in an algorithm. Geographic Information System (GIS), weather, and GPS databases could provide details such as the location, speed limit, and road condition, but do not capture the immediacy of other vehicles on the road or emergent conditions such as an animal crossing the road. One method for capturing this information is to examine the speed, lateral acceleration, and longitudinal acceleration of the vehicle. These measures capture both general road characteristics, such as the differences between driving on a state highway and through a suburban neighborhood, and also provide a direct link into these behaviors and the intentions of the driver (Fuller, 2005). Furthermore, they do not involve complex processes of database integration and can be unobtrusively collected and processed.

Like steering and pedal angle, speed and acceleration data need to be processed before they can be used as algorithm input. One such processing method is SAX, which converts continuous time-series data into discrete symbols. This conversion is accomplished through the following steps:

1 Divide the time-series into equal sized windows
2 Take the mean of the samples within each window
3 Bin the y-axis by quantiles of the normal distribution
4 Assign a letter to each quantile
5 Convert the means from step 2 into letters according to the y-axis quantiles

This process is shown in Fig. 3, which demonstrates the conversion of a continuous speed signal to the word, "dfghhhhii." SAX requires 4 sources of input: alphabet size, word length, percent overlap between windows, and window size.

In this application, SAX was applied in two ways: first to the entire 60 s windows of data leading up to each drowsiness classification and second to 5 s, 10 s, and 30 s windows of samples of speed and acceleration data. The 60 s windows captured high-level road context such as rural straight roads and highway entrance ramps. The smaller time windows captured driving maneuvers such as swerving to avoid an obstacle or overtaking another vehicle. The 60 s features used an alphabet size of 3, word length of 3, non-overlapping windows, and a window size of 60 s. These selections were made based on a sensitivity analysis of algorithm performance on training data. The shorter window features had to be aggregated across a set of smaller windows into features that represented the full 60 s. This process was accomplished by counting the frequency of each word from the smaller windows and then adding these frequency counts to a feature vector. This frequency count strategy is common in other symbolic analyses (Leslie et al., 2002). Fig. 4 illustrates the SAX micro-feature creation and algorithm integration. The features used non-overlapping windows and a word length of 3 in order to be consistent with the 60 s windows. Four different alphabet sizes were compared during algorithm selection including 3, 5, 7, and 9 letters.

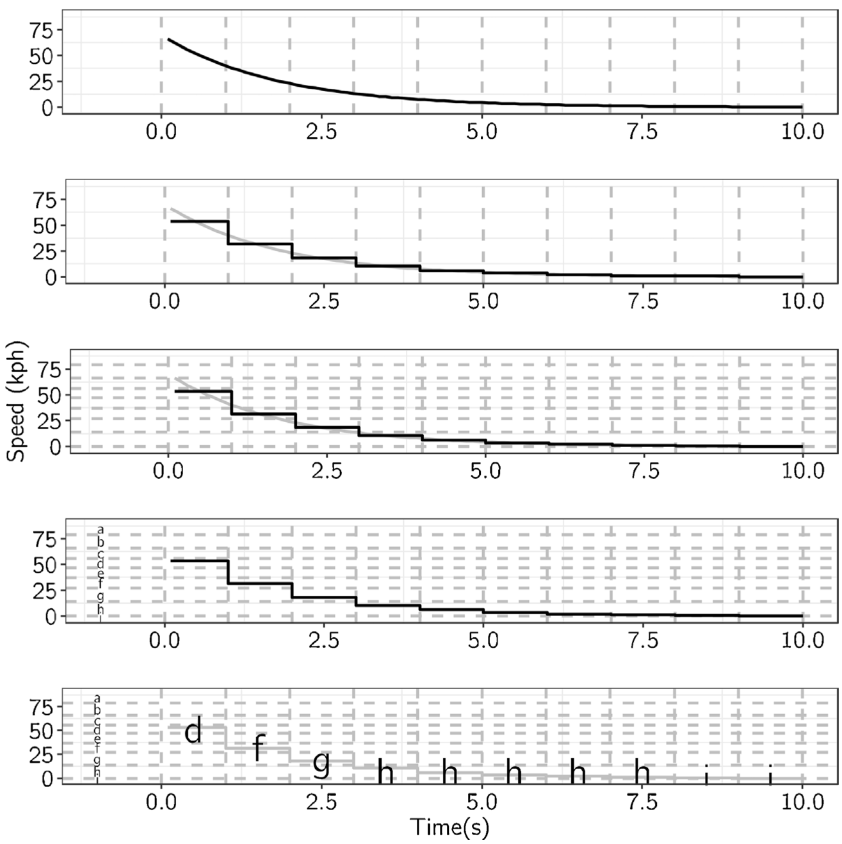

Fig. 3. Demonstration of SAX steps for a 10 s sample of data. In this example the windows are non-overlapping, window-size is set to 1 s, word length is set to 1 character, and the alphabet size is set to 9. The output of this application is the word “dfghhhhhii”.

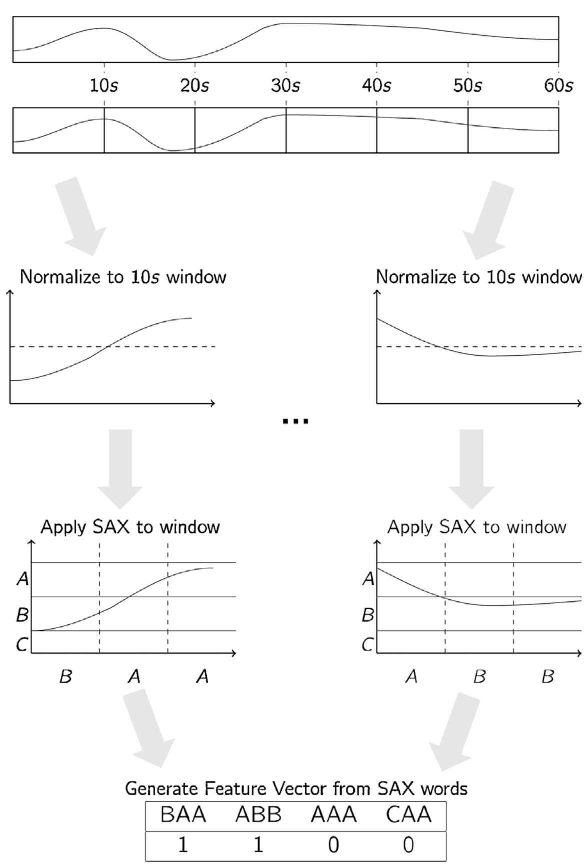

Fig. 4. Demonstration of the SAX maneuver-level feature generation process.

3. Parameter evaluation and model selection

The volume of free parameters available to DBN algorithms and SAX warrant a sensitivity analysis and model selection process. In addition to the exploration of SAX alphabet size and window size settings, this study explored four different model structures—depicted in Fig. 5. The structures differed in their inclusion of road-type context, and their treatment of the maneuver-level context. This maneuver context was treated in two ways. The first method involved including the maneuver context features with the steering and pedal features. This combination of features was used to train a random forest following the process discussed in Section 2.4.1. The trained random forest votes were used as observations or features in the DBN. This structure is depicted in the top two diagrams in Fig. 5. The second method of maneuver treatment consisted of training three separate random forest models, following the process discussed in Section 2.4.1, and incorporating them into the DBN algorithm as a multivariate normal observation. Although the difference between these structures is subtle, it is an important comparison because it tests the value of the maneuver-level context when considered independently of the pedal and steering features versus dependently. The road context inclusion and exclusion allows one to test the value of maneuver versus high-level road context and the hypothesis that with road and maneuver context combined, algorithm performance will improve on road situations that commonly involve drowsiness.

3.1. Model selection

The 4 structures and 12 SAX variations produced a total of 48 different models. The models were compared using the area under the ROC curve (AUC) and a bootstrapped confidence interval and are shown in Fig. 6. The figure shows a plot for each structure and window size pairing and four points corresponding to the various alphabet sizes investigated in the study. The model with the highest AUC: Multivariate

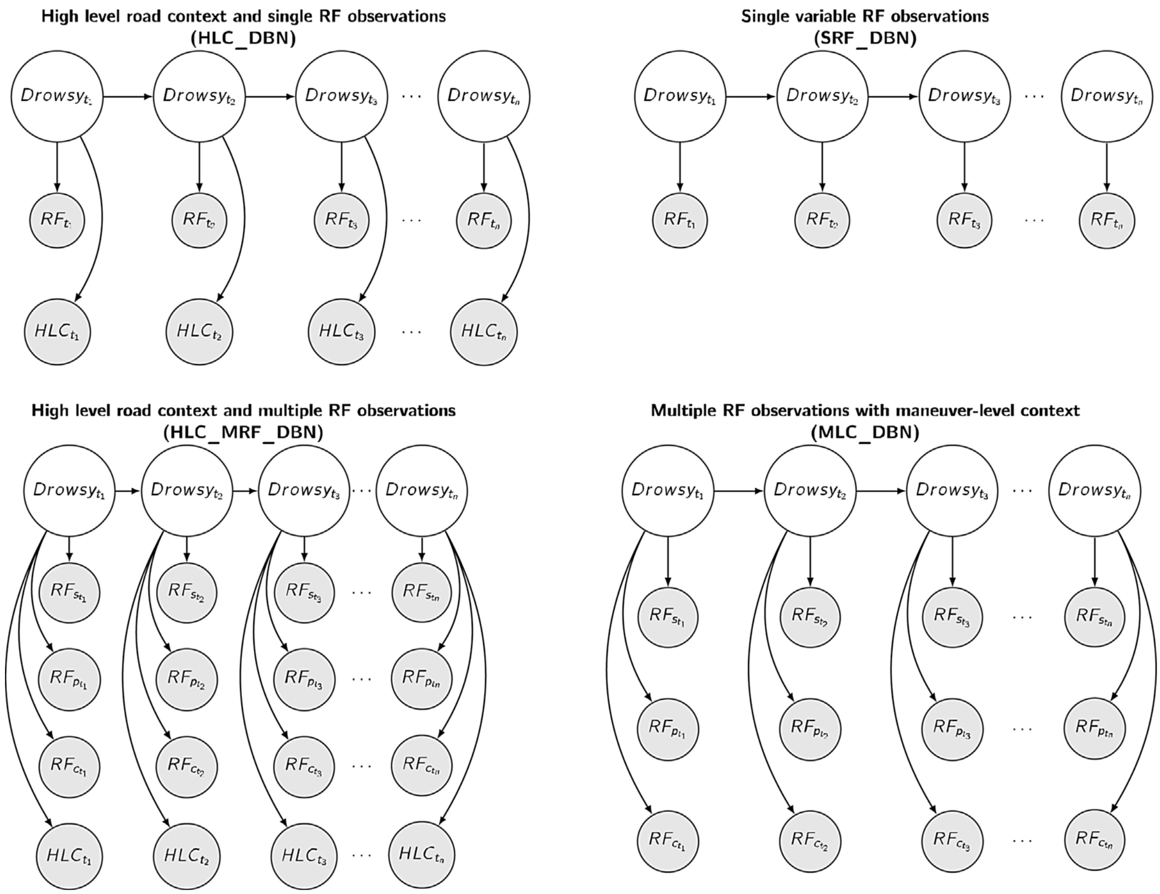

Fig. 5. Model structures examined in this study. The top and bottom rows depict structures that consolidate maneuver-level context, steering angle, and pedal input into a single random forest and structures that integrate the measures into individual random forests respectively. The left column shows structures containing both road context and maneuver-level context and the right column shows structures that contain only maneuver-level context.

random forest with no high level road context, with a window size of 10 s, and an alphabet size of 3 (MLC_DBN), is highlighted in the figure. The results suggest that an algorithm without high-level context might be preferred because it achieves the same predictive performance with fewer input variables; however, this result may reflect the particular characteristics of the sample. Future work should investigate this conclusion in more detail.

4. Results

This study examines algorithm results through comparisons to three baseline algorithms: a steering-based random forest algorithm (McDonald et al., 2013b), a DBN algorithm including steering and pedal input, and PERCLOS (Dinges and Grace, 1998). The steering-based random forest represents baseline performance of a driver-behavior based algorithm. This algorithm is a valid benchmark because it is based on steering measures but does not include temporal dependencies or context. Direct comparisons to this algorithm illustrate the effects of added complexity. The Dynamic Bayesian Network algorithm is a valuable benchmark because it includes temporal dependencies but not road context. Comparisons with this algorithm illustrate the value of adding contextual information to a prediction algorithm. PERCLOS represents an industry standard metric that has been widely employed in after-market systems (Grace et al., 1996), and will give a baseline for real-world algorithm performance. Comparisons to PERCLOS also partially illustrate each algorithm's ability to translate to real-world scenarios.

4.1. Algorithm evaluation

Fig. 7 shows the ROC curve generated by the test data predictions and a smoothed ROC curve which estimates the results on a larger test set via bootstrapped sampling. The results of the ROC plots in the left figure show that the maneuver-level context (MLC) DBN does not provide a benefit in the lower regions of the curve, where the false positive rate (FPR) is less than 0.15, but provides a substantial benefit for much of the rest of the curve, particularly over the PERCLOS (PCLS) and Steering Random Forest (RF) algorithms. The smooth ROC curve plot accentuates these differences. The AUC values for each curve were statistically evaluated relative to the maneuver-level context algorithm with a one-sided bootstrapped significance test with 2000 replicates. The replicates consisted of randomly sampled stratified samples following the method described in Carpenter and Bithell (2000). The tests found no significant differences however the difference between the AUC of the MLC DBN and PERCLOS ($D(2000) = 1.40$, $p = 0.08$), and the difference between the AUC of the MLC DBN and Steering Random Forest ($D(2000) = 1.61$, $p = 0.053$) approached significance. The gap between statistically significant findings and the observed difference in ROC curves can be reconciled by evaluating partial AUC values. Tests of the partial AUC values for false positive rates greater than 0.15 showed that the MLC DBN algorithm had significantly higher AUC than PERCLOS ($D(2000) = 1.74$, $p = 0.04$) and the Random Forest ($D(2000) = 1.86$, $p = 0.03$). Analyses at fixed thresholds provide further clarity into these findings.

Table 4 shows a comparison between the algorithms at three thresholds, one that prioritizes drowsiness detection (True Positive Rate = 1), one that balances TPR and FPR (True Positive Rate = 0.5), and one that prioritizes minimizing false positives (False Positive

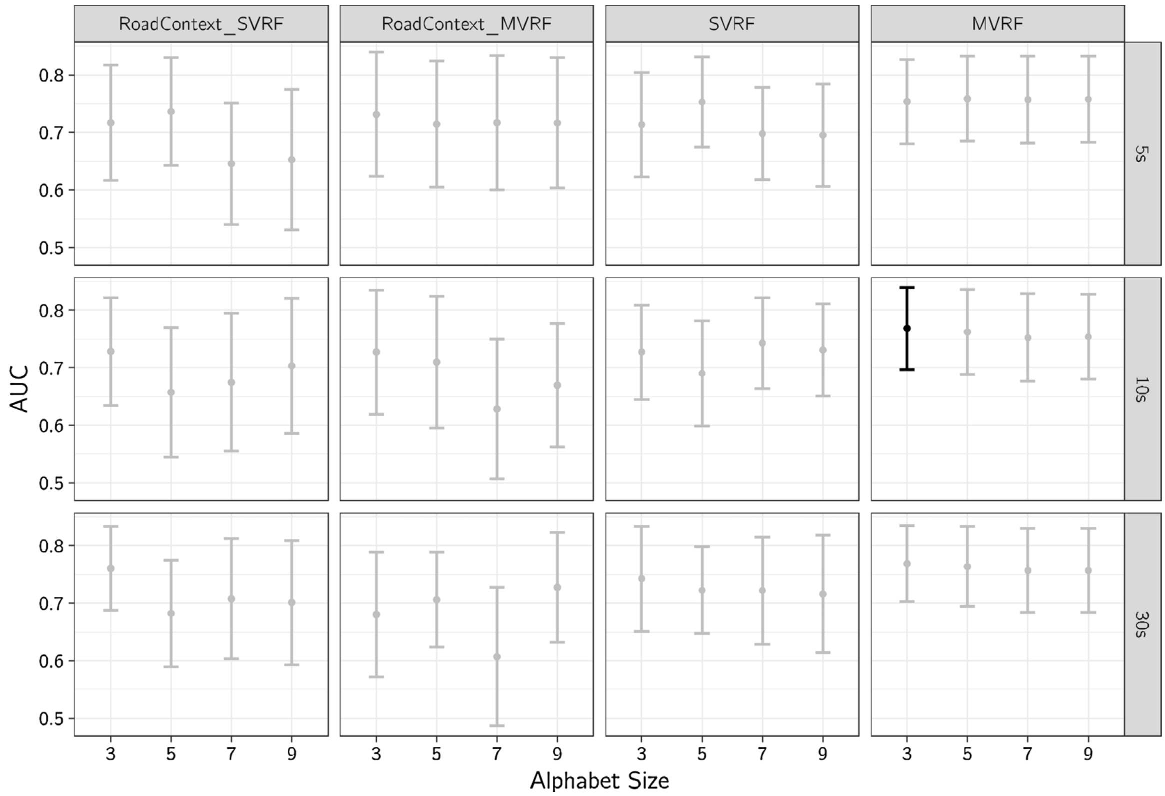

Fig. 6. AUC results with 95% bootstrapped confidence intervals for the 48 models analyzed in this study tested with the held aside test data, each plot corresponds to a model structure and window size. The structure abbreviations correspond to: Single RF with road context, Multiple RF with road context, single variable RF, and multiple variable RF. The model with the highest AUC, MVRF structure with a window size of 10 s and an alphabet size of 3, is highlighted.

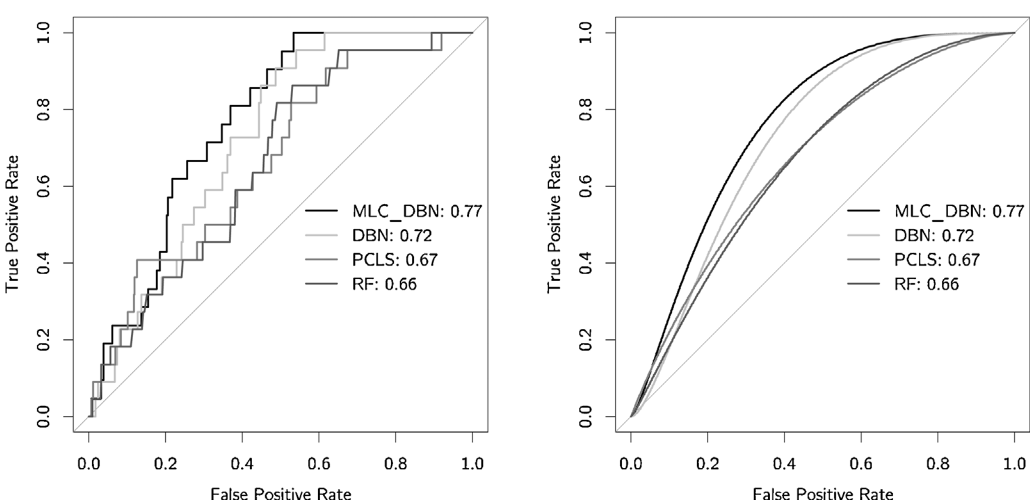

Fig. 7. ROC (left) and smoothed ROC (right) curves for the Maneuver-level context (MLC) DBN algorithm, DBN algorithm, PERCLOS, and Random Forest algorithm.

Rate = 0.1). At each threshold the table shows a confusion matrix, false positive rates, and true positive rates. The differences in true and false positive rates between the maneuver-level context DBN and the other algorithms were statistically evaluated using McNemar's test (Dietterich, 1998). For the threshold that prioritized drowsiness detection, the tests showed that the difference in false positives was significant in all cases: DBN ($M(1) = 62.35$, $p < 0.001$), PERCLOS ($M(1) = 289.80$, $p < 0.001$), Steering angle random forest ($M(1) = 224.09$, $p < 0.001$). The tests at the threshold that balanced TPR and FPR also found that the MLC DBN had significantly fewer false positives than the other algorithms: DBN ($M(1) = 50.16$, $p < 0.001$), PERCLOS ($M(1) = 69.09$, $p < 0.001$), Steering angle random forest ($M(1) = 67.54$, $p < 0.001$). The differences in TPR were not significant for the threshold that prioritized minimizing false positives.

The ROC and confusion matrix results show that adding maneuver-level context into the algorithm reduces false positives relative to algorithms that only consider time and static algorithms at thresholds that prioritize drowsiness identification or a balance FPR and TPR. However, they do not show the contexts where the algorithm improves classification performance. This improvement can be observed by analyzing false positives across contexts at a fixed threshold. False positives are a valuable metric to analyze in this case because they are unconstrained whereas the true positive rate is fixed for all algorithms. Fig. 8 shows a bar chart of false positives for each algorithm at the threshold that balances priorities across simulator environments. This threshold was selected as it is the most likely design preference. This chart shows that the maneuver-level context algorithm has fewer false positives than all other algorithms in Urban, Highway, and Rural

Table 4
Confusion matrix, TPR, and FPR for the Random Forest, Perclos, DBN, and Maneuver-level context DBN algorithms at two design preferences.

Design Preference	Algorithm	True Negatives	False Negatives	False Positives	True Positives	FPR	TPR
Prioritize drowsiness detection (TPR = 1)	RF	89	0	751	21	0.89	1
	PCLS	68	0	772	21	0.92	1
	DBN	324	0	516	21	0.61	1
	MLC DBN	392	0	448	21	0.53	1
Balanced priorities (TPR = 0.5)	RF	522	10	318	11	0.38	0.52
	PCLS	515	10	325	11	0.39	0.52
	DBN	613	10	227	11	0.27	0.52
	MLC DBN	667	10	173	11	0.20	0.52
Prioritize false positive minimization (FPR = 0.1)	RF	753	17	87	4	0.10	0.19
	PCLS	752	15	88	6	0.10	0.29
	DBN	752	16	88	5	0.10	0.24
	MLC DBN	752	16	88	5	0.10	0.24

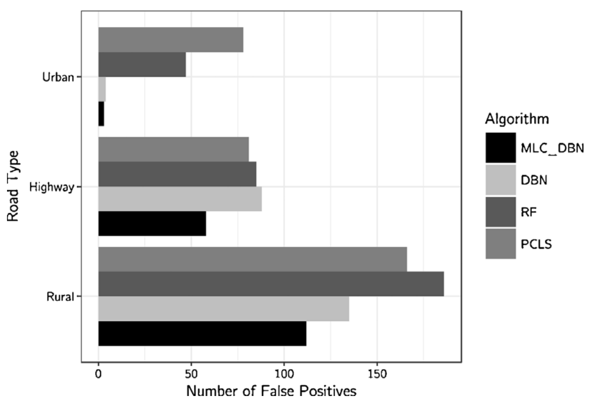

Fig. 8. Bar chart of false positives by algorithm and simulator segment. Note that the false positives were identified based on a threshold required to detect all drowsy-related lane departures.

environments. Pairwise statistical comparisons with McNemar's test showed that these differences were significant ($p < 0.001$ for all comparisons) except for the comparison with the MLC DBN and DBN algorithms in Urban Environments.

The environments in Fig. 8 can be further decomposed into the simulator events from Table 2. Fig. 9 shows such a decomposition, specifically false positives for each algorithm for a sample of simulator events. The events shown in the figure represent contexts where a drowsiness related lane departure is unlikely (Urban arterial, Interstate Curves), environments where a drowsiness lane departure is most-likely (Interstate, Rural straight drive), and environments where eye closure metrics may be compromised (Dark straight road, Dark gravel road). The figure shows that the maneuver-level context algorithm reduces false positives relative to the comparison algorithms for all events

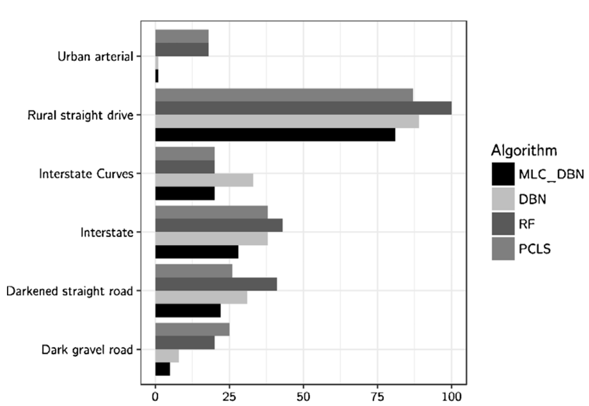

Fig. 9. Bar chart of false positive rates for each algorithm by simulator event for a sample of simulator events.

except for the Interstate Curves event. In the Interstate Curves event, the contextual algorithm has fewer false positives than the DBN algorithm but approximately the same number of false positives as the PERCLOS and the random forest algorithm.

5. Discussion

Drowsy driving crashes are a significant safety hazard that may be reduced through the use of impairment detection and warning technologies. A core component of these technologies is the detection algorithm (Balkin et al., 2011). One significant gap in current algorithms is the absence of on-road context. This study introduced a novel algorithm for drowsiness-related lane departure detection that uses vehicle input-based data, driving context features, and a temporal modeling approach to address this gap. The results of this study show that depending on the acceptable false positive rate, the context based algorithm has significantly better detection performance than simpler algorithms and PERCLOS. Specifically, the results show that an algorithm containing maneuver-level context produced better prediction performance than DBNs without context, vehicle-based algorithms that do not consider the temporal nature of driver drowsiness, and PERCLOS when drowsiness detection is prioritized over false positive rates and equivalent performance to all other algorithms when reducing the false positive rate is prioritized. Further analysis of the algorithm detection performance shows that the contextual algorithm reduces false positives across all contexts analyzed in the study including urban, highway, and rural environments. More granular analysis shows that the performance improvement is observed in contexts where drowsiness-related lane departures are likely, contexts where eye closure algorithms may be compromised, and contexts where drowsiness-related lane departures are unlikely although the effect is not persistent. This improvement partially addresses the concern that vehicle-based approaches may be sensitive to events such as turns or other common driving maneuvers (Balkin et al., 2011). These results illustrate that there is significant value in including driving context in detection algorithms.

Despite the promising results of this study, there are some counterintuitive findings in this analysis. The most prevalent of these is the finding that contextual and temporal algorithm does not significantly improve the AUC of the ROC curve relative to the benchmarks. The partial AUC analysis shows that this finding is due to the region of the curve where false positive rates are less than 15%. While these findings may be concerning, they do not invalidate the approach. These results show that for all design preferences and contexts the contextual and temporal algorithm performs at least as well as PERCLOS and the other benchmarking algorithms. With this level of performance, the contextual algorithm may be an advantageous design choice because it can be implemented without the cost or invasiveness of a camera.

Another concerning finding is the overall persistence of false positives, especially in scenarios where drowsiness-related lane departures are likely. For example, in the Rural Straight event 81 false positive predictions were observed from the MLC DBN algorithm. While this amount seems impractical it is important to understand it within the context of the ground truth definition of drowsiness employed for this study, a 1-min interval surrounding a drowsiness-related lane departure, and the binary loss function employed in the analyses. These strict definitions may negatively bias the algorithm's performance. For example, 15% of the false positives predicted by the MLC DBN occurred in adjacent 1-min time windows to drowsy instances. The current analysis treats these instances as false positives but in a practical scenario they may be viewed as early detections. Adopting this view would significantly lower the false positive rate of the algorithm. Another consideration is the predictions following a drowsiness-related lane departure. In a practical application a driver may stop driving after a warning signal or be sufficiently awaked and continue driving. In the latter case a mitigation system could be configured to reset the state of the MLC DBN to awake and then proceed with subsequent predictions as evidence is accumulated. This approach may further reduce the amount of observed false positives as the algorithm will need to gather sufficient evidence of drowsy steering and pedal input before it would classify the driver as drowsy. Beyond these improvements there is still a need for further investigation particularly because all of the benchmarking algorithms examined here showed high false positive rates. These subsequent investigations should focus on alternative definitions of drowsiness such as video coding, and larger naturalistic data sets. Broader definitions of drowsiness would support detection of the full spectrum of drowsiness from the onset of feelings of subjective drowsiness to a drowsiness-related lane departure. Large naturalistic datasets would provide larger and more diverse training and testing datasets which may improve the algorithm's performance.

The results from this study can be extended beyond drowsiness detection and mitigation to theories of drowsy driving behavior. The increase in drowsiness detection performance derived from the inclusion of maneuver-level driving context into the algorithm emphasizes the role of driving context in drowsy driving crashes and provides support for theories that consider the role of context in driver behavior. The perplexing stability of false positives in the final algorithm might be interpreted as an indication of strong similarities in driving behavior between drowsy and alert drivers on certain roads. This similarity might in turn be used to explain the lack of consistent findings relating driving behavior to drowsiness (MacLean et al., 2003; Williamson et al., 2011). Furthermore, the similarity suggests that some measure of road context should be included as a covariate in subsequent analyses of drowsy driving crashes and provides impetus for further analyses of the link between subjective drowsiness and performance decrements.

There are several limitations to this study including the use of a driving simulator, study design, the size of the test data set, and the scope of ground truth drowsiness. While this study used a high-fidelity simulator, it does not exactly replicate the experience of real driving. Furthermore, the stringent nature of the selection criteria could limit the observed effects from work schedules, pre-existing sleep disorders, and long-term poor sleep hygiene that may be observed in a naturalistic context. The inclusion criteria limit the conclusions of this study with respect to professional drivers, individuals with sleep disorders, and those that take prescription medications. Finally, while significant care was taken to differentiate between ground truth drowsiness and awake states this study did not specifically address the limitation of steering-wheel angle measures whereby steering-wheel angle patterns may be confounded between drowsiness and distraction. Future work could address these concerns by applying the methods illustrated here to a larger naturalistic driving dataset, expanding the ground truth labels in that dataset to include drowsiness, distraction, and other impairments, and combining the algorithm discussed here with the contextual factors of the Ji et al. (2006) algorithm. Future work should also address the fact that this algorithm represents a general best fit approach to all drivers. The algorithm's fit for individual drivers may be improved by synthesizing the current model and individual drivers' data through an active learning or Bayesian update approach.

6. Conclusions

This study suggests that driving context, specifically context that captures vehicle maneuvers over a 10s time window, improves detection of drowsiness-related lane departures relative to current algorithms. The results illustrate that driving context can be derived in real-time via the symbolic data reduction technique, SAX. The findings suggest that future driver drowsiness detection efforts should include driving context. These future algorithms can be combined with safety interventions targeted at causes of drowsy driving not explored in this study, such as improper work schedules and sleep disorders, to create comprehensive systems that improve driving safety.

Acknowledgement

This data collection for this study was sponsored by the National Highway Traffic Safety Administration (NHTSA), under the leadership of Eric Traube.

References

Aidman, E., Chadunow, C., Johnson, K., Reece, J., 2015. Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance. Accid. Anal. Prev. 81, 8–13. http://dx.doi.org/10.1016/j.aap.2015.03.041.

Awais, M., Badruddin, N., Drieberg, M., 2017. A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability. Sensors 17 (9), 1991. http://dx.doi.org/10.3390/s17091991.

Balkin, T.J., Horrey, W.J., Graeber, R.C., Czeisler, C.A., Dinges, D.F., 2011. The challenges and opportunities of technological approaches to fatigue management. Accid. Anal. Prev. 43 (2), 565–572. http://dx.doi.org/10.1016/j.aap.2009.12.006.

Brown, I.D., 1994. Driver fatigue. Hum. Factors J. the Hum. Factors Ergon. Soc. 36 (2), 298–314. http://dx.doi.org/10.1177/001872089403600210.

Brown, T.L., Lee, J.D., Schwarz, C., Fiorentino, D., McDonald, A.D., 2014. Assessing the feasibility of vehicle-based sensors to detect drowsy driving. Report No. DOT HS 811 886. Washington, D.C. Retrieved from. http://www.westerntransportationinstitute.org/documents/reports/4W2004_Exec_Summary.pdf.

Carpenter, J., Bithell, J., 2000. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Stat. Med. 19 (9), 1141–1164. http://dx.doi.org/10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F.

Dietterich, T., 1998. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput. 10 (7), 1895–1923. Retrieved from. http://www.ncbi.nlm.nih.gov/pubmed/9744903.

Dinges, D.F., Grace, R., 1998. PERCLOS: A Valid Psychophysiological Measure of Alertness as Assessed by Psychomotor Vigilance. Retrieved from. Federal Highway Administration, Washington D.C. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:PERCLOS+:+A+Valid+Psychophysiological+Measure+of+Alertness+As+Assessed+by+Psychomotor+Vigilance#0.

Dinges, D.F., Mallis, M.M., Maislin, G., Powell IV., J.W., 1998. Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management. Wash. DC: NHTSA.

Eskandarian, A., Mortazavi, A., 2007. Evaluation of a smart algorithm for commercial vehicle driver drowsiness detection Istanbul, Turkey: IEEE. 2007 IEEE Intelligent Vehicles Symposium 553–559. http://dx.doi.org/10.1109/IVS.2007.4290173.

Fawcett, 2004. ROC graphs: notes and practical consideration for researchers. ReCALL, 31(HPL-2003-4). pp. 1–38.

Fletcher, A., McCulloch, K., Baulk, S.D., Dawson, D., 2005. Countermeasures to driver fatigue: a review of public awareness campaigns and legal approaches. Aust. N. Z. J. Publ. Health 29 (5), 471–476. Retrieved from. http://www.ncbi.nlm.nih.gov/pubmed/16255451.

Forsman, P.M., Vila, B.J., Short, R.A., Mott, C.G., Van Dongen, H.P.A., 2013. Efficient driver drowsiness detection at moderate levels of drowsiness. Accid. Anal. Prev. 50 (null), 341–350. http://dx.doi.org/10.1016/j.aap.2012.05.005.

Fuller, R., 2005. Towards a general theory of driver behaviour. Accid. Anal. Prev. 37 (3), 461–472. http://dx.doi.org/10.1016/j.aap.2004.11.003.

Furman, G.D., Baharav, A., Cahan, C., Akselrod, S., 2008. Early detection of falling asleep at the wheel: a heart rate variability approach IEEE. 2008 Computers in Cardiology 1109–1112. http://dx.doi.org/10.1109/CIC.2008.4749240.

Gander, P., Hartley, L., Powell, D., Cabon, P., Hitchcock, E., Mills, A., Popkin, S., 2011. Fatigue risk management: organizational factors at the regulatory and industry/company level. Accid. Anal. Prev. 43 (2), 573–590. http://dx.doi.org/10.1016/j.aap.2009.11.007.

Garcés Correa, A., Orosco, L., Laciar, E., 2014. Automatic detection of drowsiness in EEG records based on multimodal analysis. Med. Eng. Phys. 36 (2), 244–249. http://dx.doi.org/10.1016/j.medengphy.2013.07.011.

Geist, G.F., Merkt, R.,Altamuro S. N.J.S.2C:11-5, 4 § (2002). State of New Jersey.

Ghazizadeh, M., Peng, Y., Lee, J.D., Boyle, L.N., 2012. Augmenting the technology acceptance model with trust: commercial drivers' attitudes towards monitoring and feedback. Proc. Hum. Factors Ergon. Soc. Annu. Meet 56 (1), 2286–2290. http://dx.doi.org/10.1177/1071181312561481.
Grace, R., Byrne, V.E., Bierman, D.M., Legrand, J.-M., Gricourt, D., Davis, B.K., et al., 1996. A drowsy driver detection system for heavy vehicles IEEE. 17th DASC. AIAA/IEEE/SAE. Digital Avionics Systems Conference. Proceedings (Cat. No.98CH36267) 2http://dx.doi.org/10.1109/DASC.1998.739878. pp. I36/1–I36/8.
Grace, R., Stewart, S., 2001. The Co-pilot: a low-cost drowsy driver and driver inattention monitor Aspen, CO. The International Driving Symposium on Human Factors in Driver Assessment, Training And Vehicle Design.
Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182. Retrieved from. http://dl.acm.org/citation.cfm?id=944968.
Hanowski, R.J., Bowman, D., Alden, A., Wierwille, W.W., Carroll, R., 2008a. PERCLOS+: development of a robust field measure of driver drowsiness New York. 15th World Congress on Intelligent Transport Systems and ITS America's 2008 Annual Meeting.
Hanowski, R.J., Bowman, D.S., Wierwille, W.W., Alden, A., Carroll, R., 2008b. PERCLOS+: moving beyond single metric systems. Proceedings of the Society of Automotive Engineers Commercial Vehicle Engineering Conference.
Hartley, L., Horberry, T., Mabbott, N., 2000. Review of Fatigue Detection and Prediction Technologies. Melbourne.
Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., Dement, W.C., 1973. Quantification of sleepiness: a new approach. Psychophysiology 10 (4), 431–436. http://dx.doi.org/10.1111/j.1469-8986.1973.tb00801.x.
Hu, S., Zheng, G., 2009. Driver drowsiness detection with eyelid related parameters by support vector machine. Expert Syst. Appl. 36 (4), 7651–7658. http://dx.doi.org/10.1016/j.eswa.2008.09.030.
Ji, Q., Lan, P., Looney, C., 2006. A probabilistic framework for modeling and real-time monitoring human fatigue. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 36 (5), 862–875. http://dx.doi.org/10.1109/TSMCA.2005.855922.
Ji, Q., Zhu, Z., Lan, P., 2004. Real-time nonintrusive monitoring and prediction of driver fatigue. IEEE Trans. Veh. Technol. 53 (4), 1052–1068. http://dx.doi.org/10.1109/TVT.2004.830974.
Jo, J., Lee, S.J., Park, K.R., Kim, I.J., Kim, J., 2014. Detecting driver drowsiness using feature-level fusion and user-specific classification. Expert Syst. Appl. 41 (4 Part 1), 1139–1152. http://dx.doi.org/10.1016/j.eswa.2013.07.108.
Klauer, S., Dingus, T., Neale, V., 2006. The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. Washington D.C. Retrieved from. http://trid.trb.org/view.aspx?id=786825.
Kotsiantis, S.B., Zaharakis, I.D., Pintelas, P.E., 2007. Supervised machine learning : a review of classification techniques. Informatica 31 (3), 501–520.
Krajewski, J., Golz, M., Sommer, D., 2009. Detecting sleepy drivers by pattern recognition based analysis of steering wheel behaviour. Der Mensch Im Mittelpunkt Technischer Systeme. pp. 285–288. Retrieved from. https://www.tu-berlin.de/fileadmin/f25/dokumente/8BWMMS/15.5-Krajewski.pdf.
Krajewski, J., Sommer, D., 2009. Steering wheel behavior based estimation of fatigue. Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment and Design. pp. 118–124. Retrieved from. http://www.researchgate.net/publication/228688504_Steering_Wheel_Behavior_Based_Estimation_of_Fatigue/file/79e4150a91cc037eae.pdf.
Kutila, M.H., Jokela, M., Mäkinen, T., Viitanen, J., Markkula, G., Victor, T.W., 2007. Driver cognitive distraction detection: feature estimation and implementation. Proc. Inst. Mech. Eng. Part D J. Automob. Eng. 221 (9), 1027–1040. http://dx.doi.org/10.1243/09544070JAUTO332.
Lal, S.K.L., Craig, A., 2001. A critical review of the psychophysiology of driver fatigue. Biol. Psychol. 55 (3), 173–194. http://dx.doi.org/10.1016/S0301-0511(00)00085-5.
Lal, S.K.L., Craig, A., Boord, P., Kirkup, L., Nguyen, H., 2003. Development of an algorithm for an EEG-based driver fatigue countermeasure. J. Saf. Res. 34 (3), 321–328. http://dx.doi.org/10.1016/S0022-4375(03)00027-6.
Lee, B.G., Lee, B.L., Chung, W.Y., 2015. Smartwatch-based driver alertness monitoring with wearable motion and physiological sensor. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2015–Novem 6126–6129. http://dx.doi.org/10.1109/EMBC.2015.7319790.
Lee, J.D., See, K.A., 2004. Trust in automation: designing for appropriate reliance. Hum. Factors J. Hum. Factors Ergon. Soc. 46 (1), 50–80.
Lenné, M.G., Jacobs, E.E., 2016. Predicting drowsiness-related driving events: a review of recent research methods and future opportunities. Theor. Issues Ergon. Sci. 17 (5–6), 533–553. http://dx.doi.org/10.1080/1463922X.2016.1155239.
Leslie, C., Eskin, E., Noble, W.S., 2002. The spectrum kernel: a string kernel for SVM protein classification Lihue, Hawaii. Pacific Symposium on Biocomputing 7, 566–575.
Lin, J., Keogh, E., Wei, L., Lonardi, S., 2007. Experiencing SAX: a novel symbolic representation of time series. Data Min. Knowl. Discov. 15 (2), 107–144. http://dx.doi.org/10.1007/s10618-007-0064-z.
Liu, C.C., Hosking, S.G., Lenné, M.G., 2009. Predicting driver drowsiness using vehicle measures: recent insights and future challenges. J. Saf. Res. 40 (4), 239–245. http://dx.doi.org/10.1016/j.jsr.2009.04.005.
MacLean, A.W., Davies, D.R., Thiele, K., 2003. The hazards and prevention of driving while sleepy. Sleep Med. Rev. 7 (6), 507–521. http://dx.doi.org/10.1016/S1087-0792(03)90004-9.
Maycock, 1997. Sleepiness and driving: the experience of U.K. car drivers. Accid. Anal. Prev. 29 (4), 453–462.
McDonald, A.D., Lee, J.D., Aksan, N.S., Dawson, J.D., Tippin, J., Rizzo, M., 2013a. The language of driving. Transp. Res. Rec. J. Transp. Res. Board 2392 (1), 22–30. http://dx.doi.org/10.3141/2392-03.
McDonald, A.D., Lee, J.D., Aksan, N.S., Dawson, J.D., Tippin, J., Rizzo, M., 2017. Using kinematic driving data to detect sleep apnea treatment adherence. J. Intell. Transp. Syst. 2450 (September), 1–13. http://dx.doi.org/10.1080/15472450.2017.1369060.
McDonald, A.D., Lee, J.D., Schwarz, C., Brown, T.L., 2013b. Steering in a random forest: ensemble learning for detecting drowsiness-related lane departures. Hum. Factors J. Hum. Factors Ergon. Soc. http://dx.doi.org/10.1177/0018720813515272.
McRuer, D., Allen, R., Weir, D., Klein, R., 1977. New results in driver steering control models. Hum. Factors J. Hum. Factors Ergon. Soc. 19 (4), 381–397. http://dx.doi.org/10.1177/001872087701900406.
Michon, J.A., 1986. A critical view of driver behavior models: what do we know, what should we do? Human Behavior and Traffic Safety. Springer, US, pp. 485–524. http://dx.doi.org/10.1007/978-1-4613-2173-6_19.
Murata, A., 2016. Proposal of a method to predict subjective rating on drowsiness using physiological and behavioral measures. IIE Trans. Occup. Ergon. Hum. Factors 7323 (March). http://dx.doi.org/10.1080/21577323.2016.1164765.
Murphy, K.P., 2002. Dynamic bayesian networks. In: Jordan, M. (Ed.), Probabilistic Graphical Models, Retrieved from. http://webdocs.cs.ualberta.ca/~greiner/C-366/366-SLIDES/dbn-murphy.pdf.
NADS, 2010. National Advanced Driving Simulator. Retrieved from. http://www.nads-sc.uiowa.edu/sim_nads1.php.
NHTSA, 2011. Traffic Safety Facts: Drowsy Driving. Washington D.C. Retrieved from. http://www-nrd.nhtsa.dot.gov/pubs/811449.pdf.
Pack, A.I., Pack, A.M., Rodgman, E., Cucchiara, A., Dinges, D.F., Schwab, C.W., 1995. Characteristics of crashes attributed to the driver having fallen asleep. Accid. Anal. Prev. 27 (6), 769–775. http://dx.doi.org/10.1016/0001-4575(95)00034-8.
Patel, M., Lal, S.K.L., Kavanagh, D., Rossiter, P., 2011. Applying neural network analysis on heart rate variability data to assess driver fatigue. Expert Syst. Appl. 38 (6), 7235–7242. http://dx.doi.org/10.1016/j.eswa.2010.12.028.
Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., Müller, M., 2011. pROC: an open-source package for R and S++ to analyze and compare ROC curves. BMC Bioinform. 12 (Swets 1973), 77. http://dx.doi.org/10.1186/1471-2105-12-77.
Sagberg, F., 1999. Road accidents caused by drivers falling asleep. Accid. Anal. Prev. 31 (6), 639–649. Retrieved from. http://www.ncbi.nlm.nih.gov/pubmed/10487339.
Salvucci, D., 2006. Modeling driver behavior in a cognitive architecture. Hum. Factors J. Hum. Factors Ergon. Soc. 48 (2), 362–380. http://dx.doi.org/10.1518/001872006777724417.
Sandberg, D., Akerstedt, T., Anund, A., Kecklund, G., Wahde, M., 2011. Detecting driver sleepiness using optimized nonlinear combinations of sleepiness indicators. IEEE Trans. Intell. Transp. Syst. 12 (1), 97–108. http://dx.doi.org/10.1109/TITS.2010.2077281.
Sayed, R., Eskandarian, A., 2001. Unobtrusive drowsiness detection by neural network learning of driver steering. Proc. Inst. Mech. Eng. Part D J. Automob. Eng. 215 (9), 969–975. http://dx.doi.org/10.1243/0954407011528536.
Sun, W., Zhang, X., Peeta, S., He, X., Li, Y., 2017. A real-time fatigue driving recognition method incorporating contextual features and two fusion levels. Transportation Research Board, 96th Annual Meeting. pp. 1–13.
Tijerina, L., Gleckler, M., Stoltzfus, D., 1999. A preliminary assessment of algorithms for drowsy and inattentive driver detection on the road, vol. 808 Washington D.C. Retrieved from. http://trid.trb.org/view.aspx?id=502406.
Van Loon, R.J., Brouwer, R.F.T., Martens, M.H., 2015. Drowsy drivers' under-performance in lateral control: how much is too much? Using an integrated measure of lateral control to quantify safe lateral driving. Accid. Anal. Prev. 84, 134–143. http://dx.doi.org/10.1016/j.aap.2015.08.012.
Vuckovic, A., Radivojevic, V., Chen, A.C.N., Popovic, D., 2002. Automatic recognition of alertness and drowsiness from EEG by an artificial neural network. Med. Eng. Phys. 24 (5), 349–360. http://dx.doi.org/10.1016/S1350-4533(02)00030-9.
Wang, J., Sun, S., Fang, S., Fu, T., Stipancic, J., 2017. Predicting drowsy driving in real-time situations: using an advanced driving simulator, accelerated failure time model, and virtual location-based services. Accid. Anal. Prev. 99, 321–329. http://dx.doi.org/10.1016/j.aap.2016.12.014.
Wang, M.S., Jeong, N.T., Kim, K.S., Choi, S.B., Yang, S.M., You, S.H., et al., 2016. Drowsy behavior detection based on driving information. Int. J. Automot. Technol. 17 (1), 165–173. http://dx.doi.org/10.1007/s12239-016-0016-y.
Wang, Q., Yang, J., Ren, M., Zheng, Y., 2006. Driver fatigue detection: a survey Dalian, China: IEEE. 2006 6th World Congress on Intelligent Control and Automation 8587–8591. http://dx.doi.org/10.1109/WCICA.2006.1713656.
Weir, D., McRuer, D., 1970. Dynamics of driver vehicle steering control. Automatica 6 (1), 87–98. Retrieved from. http://www.sciencedirect.com/science/article/pii/0005109870900774.
Wierwille, W.W., Ellsworth, L.A., Wreggit, S.S., Fairbanks, R.J., Kim, C.L., 1994a. Research on Vehicle-Based Driver Status/Performance Monitoring: Development, Validation, and Refinement of Algorithms for Detection of Driver Drowsiness. National Highway Traffic Safety Administration, Washington DC.
Wierwille, W.W., Wreggit, S.S., Kirn, C.L., Ellsworth, L.A., Fairbanks, R.J., 1994b. Research on Vehicle-Based Driver Status/Performance Monitoring: Development, Validation, and Refinement of Algorithms for Detecting Driver Drowsiness. Washington D.C. Retrieved from. http://trid.trb.org/view.aspx?id=448128.
Wilde, G.J.S., 1988. Risk homeostasis theory and traffic accidents: propositions, deductions and discussion of dissension in recent reactions. Ergonomics 31 (4), 441–468. http://dx.doi.org/10.1080/00140138808966691.
Wilkinson, R.T., Houghton, D., 1982. Field test of arousal: a portable reaction timer with data storage. Hum. Factors J. Hum. Factors Ergon. Soc. 24 (4), 487–493. http://dx.doi.org/10.1177/001872088202400409.
Williamson, A., Lombardi, D.A., Folkard, S., Stutts, J., Courtney, T.K., Connor, J.L., 2011. The link between fatigue and safety. Accid. Anal. Prev. 43 (2), 498–515. http://dx.doi.org/10.1016/j.aap.2009.11.011.

Wilson, B.J., Bracewell, T.D., 2000. Alertness monitor using neural networks for EEG analysis Sydney, NSW: IEEE. Neural Networks for Signal Processing X. Proceedings of the 2000 IEEE Signal Processing Society Workshop (Cat. No.00TH8501) 2, 814–820. http://dx.doi.org/10.1109/NNSP.2000.890161.

Yang, G., Lin, Y., Bhattacharya, P., 2010. A driver fatigue recognition model based on information fusion and dynamic Bayesian network. Inf. Sci. 180 (10), 1942–1954.

Yang, J.H., Tijerina, L., Pilutti, T., Coughlin, J.F., Feron, E., 2009. Detection of driver fatigue caused by sleep deprivation. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 39 (4), 694–705. http://dx.doi.org/10.1109/TSMCA.2009.2018634.

Yeo, M.V.M., Li, X., Shen, K., Wilder-Smith, E.P.V., 2009. Can SVM be used for automatic EEG detection of drowsiness during car driving? Saf. Sci. 47 (1), 115–124. http://dx.doi.org/10.1016/j.ssci.2008.01.007.

Zhao, C., Zhao, M., Liu, J., Zheng, C., 2012. Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. Accid. Anal. Prev. 45, 83–90. http://dx.doi.org/10.1016/j.aap.2011.11.019.

Zilberg, E., Xu, Z.M., Burton, D., Karrar, M., Lal, S.K.L., 2007. Methodology and initial analysis results for development of non-invasive and hybrid driver drowsiness detection systems Sydney, NSW: IEEE. The 2nd International Conference on Wireless Broadband and Ultra Wideband Communications (AusWireless 2007). pp. 16. http://dx.doi.org/10.1109/AUSWIRELESS.2007.44.