International Journal of Mobile Human Computer Interaction
Volume 11 • Issue 3 • July-September 2019

Using Naturalistic Vehicle-Based Data to Predict Distraction and Environmental Demand

Dina Kanaan, University of Toronto, Toronto, Canada
Suzan Ayas, University of Toronto, Toronto, Canada
Birsen Donmez, University of Toronto, Toronto, Canada
Martina Risteska, University of Toronto, Toronto, Canada
Joyita Chakraborty, University of Toronto, Toronto, Canada

ABSTRACT

This research utilized vehicle-based measures from a naturalistic driving dataset to detect distraction as indicated by long off-path glances ($\geq$ 2 s) and whether the driver was engaged in a secondary (non-driving) task or not, as well as to estimate motor control difficulty associated with the driving environment (i.e. curvature and poor surface conditions). Advanced driver assistance systems can exploit such driver behavior models to better support the driver and improve safety. Given the temporal nature of vehicle-based measures, Hidden Markov Models (HMMs) were utilized; GPS speed and steering wheel position were used to classify the existence of off-path glances (yes vs. no) and secondary task engagement (yes vs. no); lateral (x-axis) and longitudinal (y-axis) acceleration were used to classify motor control difficulty (lower vs. higher). Best classification accuracies were achieved for identifying cases of long off-path glances and secondary task engagement with both accuracies of 77%.

KEYWORDS

Driver State Detection, Glance Behavior, Hidden Markov Models, Naturalistic Driving Study, NEST, Off-Path Glances, Secondary Task Engagement, SHRP2

INTRODUCTION

Driving is a complex task that mainly requires visual perception and manual control. The performance of this task is influenced by the driver's ability to perceive and adapt to different environmental demands, which in turn is influenced by driver's state. Distraction, a driver state that is defined as "the diversion of attention away from activities critical for safe driving toward a competing activity" by Regan, Lee, and Young (2008, p. 34), has been shown to be prevalent among drivers (Dingus et al., 2016; Young & Lenné, 2010), to impair driving performance (Caird, Willness, Steel, & Scialfa, 2008; Horrey & Wickens, 2006; Regan et al., 2008), and to increase crash risk (Dingus et al., 2016; National Highway Traffic Safety Administration, 2018).

DOI: 10.4018/IJMHCI.2019070104

Technological advances are generating smarter vehicles that aim to enhance traffic safety and the experience of driving by detecting driver state (McCall & Trivedi, 2004; Pentland & Liu, 1999; Yang, Lin, & Bhattacharya, 2010), the state of the driving environment (Fridman et al., 2016), and driver intent (Jain, Koppula, Raghavan, Soh, & Saxena, 2015; Martin, Vora, Yuen, & Trivedi, 2018). These technologies that are mainly still in development are promising for distraction mitigation. For example, upon predicting driving impairment due to distraction, the vehicle may provide feedback to the driver that could help direct the driver's attention back to the driving task (Donmez, Boyle, & Lee, 2007, 2008) or may take over lateral and/or longitudinal control (e.g. Stanton & Young, 2005). Further, when high levels of driver distraction are detected, the vehicle can adapt user interfaces that are built-in or brought into the vehicle, such as by filtering information content, delaying notifications, and blocking access to certain actions. For example, Tchankue, Wesson, and Vogts (2011) developed a prototype adaptive user interface for an in-car communication system that blocked incoming phone calls and prevented drivers from sending text messages when distraction was detected. Such a system can also block phone calls when environmental demands are predicted to be high. In general, better system awareness of the driving environment can lead to more effective and intelligent distraction mitigation strategies.

Driver Distraction and Environmental Demand Detection

Several sources of information can be used for driver state detection. Aghaei et al. (2016) classified driver state detection measures into three categories: vehicle-based, physiological, and facial or body expression. Vehicle-based measures can be directly related to driver input (e.g. steering and braking), the vehicle's response to driver input (e.g. velocity and acceleration), and the vehicle's state relative to the environment (e.g. headway distance and lane deviations). Physiological measures are indicators of physical arousal, and include measures such as heart rate, skin conductance, and brain activity. Facial or body expression measures involve techniques like face detection, eye tracking, and motion detection, and provide measures such as secondary (i.e. non-driving related) task engagement and off-path glance durations. Extensive research has documented the reactivity of these different measures to driver distraction (see Aghaei et al. (2016) for a review), although some measures assess distraction more directly than others. For example, increased heart rate may be due to distraction, but it could also be due to stress (Taelman, Vandeput, Spaepen, & Van Huffel, 2009) or an increase in the ambient temperature (Achten & Jeukendrup, 2003). On the other hand, off-road glances would unambiguously indicate that the driver's visual attention is not allocated to the roadway. In fact, there is a strong established relation between long off-road glances, in particular, glances longer than 2 seconds, and increased crash risk (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006). Secondary task engagement, which can be inferred from observations of body movement, can be regarded as another proxy for driver distraction, as it is a strong indicator that the driver's attention is directed toward another activity that may compete with driving and introduce safety risks.

Although both physiological and facial or body expression measures can be useful in distraction detection, these measures generally require cameras/computer vision and sophisticated physiological sensors. While these technologies work well in driving simulator studies, there are concerns about the feasibility and reliability of employing such methods in the real world, particularly in production vehicles. Cameras, eye-tracking technology, and physiological sensors are known to be affected by environmental factors such as motion and lighting, are costly to implement, and can be intrusive to the driver. Moreover, the use of these technologies to record driver face and physiological activity may bring about concerns regarding privacy and data confidentiality. Vehicle-based measures, on the other hand, especially kinematic measures such as velocity and acceleration, and driver input measures such as steering wheel position, are readily accessible from sensors that are already built into vehicles. Thus, these vehicle-based measures provide a good starting point for attempting to detect driver state in the real world.

There has been a variety of attempts to detect driver distraction. However, most research has leveraged facial or body expression measures (e.g. Lee et al., 2018), whereas few have focused solely on vehicle-based measures (e.g. Atiquzzaman, Qi, & Fries, 2018; Tango & Botta, 2013). Further, some of these studies used simulator (e.g. Atiquzzaman, Qi, & Fries, 2018; Tango & Botta, 2013) or instrumented vehicle data (Kircher & Ahlstrom, 2010; Li, Bao, Kolmanovsky, & Yin, 2018). As discussed above, the equipment reliability issues are not as prominent in the simulator as they are in the real world. In addition, simulator and instrumented vehicle studies employ a degree of experimental control and therefore may not truly represent how drivers are distracted in the real world. For the latter reason, naturalistic driving studies have increasingly been performed to understand drivers' natural behavior on the road, including their interaction with in-vehicle systems and other sources of distraction. Naturalistic approaches make use of cameras and sensors installed on participating drivers' vehicles that collect data for extended periods of time. For example, the Second Strategic Highway Research Program 2 Naturalistic Driving Study (SHRP2 NDS) – the largest naturalistic driving study conducted in the United States so far – contains data collected over a three-year period from over 3000 vehicles and around 50 million miles driven (Dingus et al., 2014). Naturalistic study datasets, such as SHRP2, provide a great opportunity for evaluating the feasibility of driver state detection in the real world from vehicle-based measures.

Li et al. (2018) utilized instrumented vehicles to collect driving data from 16 drivers and classified visual-manual secondary task engagement using vehicle-based measures including steering wheel entropy as well as the standard deviation of steering wheel position, of speed, and of headway distance. They were able to use headway distance as a predictor given that they collected data via instrumented vehicles and reached high overall levels of accuracy (~95%) but high levels of false positive rates (~78%). Their models were trained on the individual driver data given that they only had 16 drivers. Kircher and Ahlstrom (2010) used vehicle-based measures such as speed, speed variability, throttle hold rate, and steering wheel reversal rate to predict instances of "distraction" labeled based on non-driving related glances. They achieved an accuracy rate of 76% but their data was collected in a small-scale instrumented vehicle study with seven drivers. In general, further attempts are needed to investigate driver distraction detection via vehicle-based measures with larger sample sizes and in more naturalistic settings.

In addition to inferring driver distraction, vehicle-based measures may also be leveraged for detecting environmental demand, which can lead to smarter distraction mitigation strategies. For example, crash severity is higher on curves (Donmez & Liu, 2015) and crash risk increases with road slipperiness (Norrman, Eriksson, & Lindqvist, 2000), which indicate that road alignment and road surface conditions affect motor control demands experienced by drivers.

Objective

In this paper, naturalistic driving data, in particular, the Naturalistic Engagement in Secondary Task (NEST) dataset, was utilized to investigate whether vehicle-based measures can be exploited to predict long off-path glances and secondary task engagement (indicators of distraction), as well as motor control difficulty (an indicator of environmental demand). The NEST dataset is a sample of the SHRP2 data that was created for the purpose of studying driver distraction in a naturalistic setting (Owens, Angell, Hankey, Foley, & Ebe, 2015). For prediction, Hidden Markov Models (HMMs) were utilized because they have been found to be successful at predicting time-sequential data (Yamato, Ohya, & Ishii, 1992) and were found to generate high accuracies in driver state prediction, e.g. Lee et al. (2018).

METHODS

Using the NEST data, HMMs were built from vehicle-based measures (in particular, GPS speed, steering wheel position, and lateral and longitudinal acceleration) to predict the existence of long off-path glances, secondary task engagement, and higher levels of motor control difficulty. Separate models were built for each variable.

NEST Dataset and Feature Selection

The NEST dataset contains baseline data (i.e. data from normal conditions with no crashes or near-crashes) from 204 drivers. Overall, the dataset consists of a total of 944 baselines, each 20 seconds long and broken into two 10-second long epochs for coding purposes. NEST variables used as predictors in our hidden Markov models are summarized in Table 1, while the NEST variables that were used for creating the categories of long off-path glances (yes vs. no), secondary task engagement (yes vs. no), and motor control difficulty (higher vs. lower) are summarized in Table 2. A detailed listing and description of all NEST variables can be found in Owens et al. (2015). Sequential variables listed in these two tables are reported at a rate of 1 (GPS speed) or 10 (steering wheel angle, acceleration, and gaze direction) frames per second, which result in around 20 or 200 samples for each 20-second-long baseline, while aggregated variables are reported in a manner that represents a high-level summary of the state of each variable (secondary task engagement, road surface condition, and road alignment) during each 10-second epoch. For example, if there is curvature detected in an epoch, the alignment of the road would be labeled as "curved" for the entire 10 second epoch.

GPS speed, steering wheel position, and lateral and longitudinal acceleration were chosen as predictors for our models because previous studies have found steering wheel and speed measures useful in predicting distraction (Kircher & Ahlstrom, 2010; Li et al., 2018), and lateral and longitudinal acceleration are expected to be affected by motor control difficulty. Thus,

Table 1. Summary of NEST variables used as predictors in HMMs

Variable	Source and Type	Unit	Sampling Rate	Description
GPS Speed	Vehicle kinematics / Sequential	km/h	1 Hz (~20 samples per baseline)	Speed measured by GPS
Steering Wheel Position		degrees	10 Hz (~200 samples per baseline)	Steering wheel angle
x-acceleration		g		Longitudinal acceleration
y-acceleration		g		Lateral acceleration

Table 2. Summary of NEST variables used for categorizing output variables of HMMs

Variable	Source	Type	Sampling Rate	Description
Gaze Direction	Annotated video sequences	Sequential	10 Hz (~ 200 samples per baseline event)	Allocation of drivers' gaze to different areas such as forward (on-path), center stack, right mirror, and left mirror, as well as the transitions between areas
Secondary Task		Aggregated	Once for 10 seconds (2 samples per baseline)	Up to six out of 40 different task types: e.g. interaction with carried-in devices, grooming, interaction with passengers
Surface Condition	Environment variables			Good vs. poor (good if the road surface is dry for both 10-second epochs; else, poor)
Road Alignment				Straight vs. curved (straight if the road is straight for both 10-second epochs; else, curved)

GPS speed and steering wheel position were used for classifying the existence of long-off path glances and secondary task engagement, whereas lateral and longitudinal acceleration were used to classify motor control difficulty.

Predicted Classes

Table 3 defines the classes (or levels) of the three predicted variables (i.e. long off-path glance, secondary task engagement, and motor control difficulty) and presents the NEST variables used in creating these classes. Further details are provided in the following sections.

Long Off-Path Glance Classes

The NEST dataset includes gaze position categorized into 18 possible areas or locations of interest (e.g. forward or on-path, center stack, and left window/mirror). The dataset also reports gaze transitions between different areas. Baselines with missing values for gaze position or with a value of "unknown" (n = 36) were excluded from our analysis, leaving 908 baselines that included 190,535 frames with gaze position information. These frames were used to identify whether there were any off-path glances (i.e. non-forward) that were 2 seconds or longer in a given baseline. A threshold of 2 seconds was used given that glances longer than 2 seconds have been found to significantly increase crash risk (Klauer et al., 2006). Overall, 182 baselines were labeled as having at least one long off-path glance, the remaining 726 baselines were labeled as not having long off-path glances. When calculating off-path glance durations, the transition time preceding the off-path glance was included, according to the ISO 16673:2007(E) standard (International Organization for Standardization, 2007).

Secondary Task Engagement Classes

The NEST dataset reports on 40 different categories of secondary tasks (e.g. texting, cellphone related engagements, dancing, singing, interacting with passengers, and grooming). If a baseline had a record of any of these secondary tasks, then it was labeled as having secondary task engagement. All other baselines were categorized as having no secondary task engagement. Although NEST categorizes the presence of passengers/children with no interaction as secondary tasks, we labeled these cases as "no secondary task engagement". Baselines that had "unknown" or "N/A" secondary tasks were excluded from the analysis, leaving 928 baselines: 701 with secondary task engagement and 227 with

Table 3. Predicted variables and their classes

Predicted Variable	Nest Variable(s)	Classes	Number of Baselines	Number of Baselines Used in HMMs
Long Off-path Glances (≥ 2 s)	Gaze Direction	**Yes:** Baseline contains at least one off-path glance that is 2 seconds or longer	182	34
		No: Baseline contains no off-path glances 2 seconds or longer	726	143
Secondary Task Engagement	Secondary Task	**Yes:** Driver engaged in at least one secondary task	701	133
		No: Driver not engaged in any secondary tasks	227	46
Motor Control Difficulty	Surface Condition, Road Alignment	**Higher:** Either poor surface condition, curved road, or both	315	275
		Lower: Good (dry) surface condition and straight road	617	538

no secondary task engagement. Overall, the 928 baselines corresponded to a total of 186,192 frames of vehicle-based data recorded at 10 Hz.

Motor Control Difficulty Classes

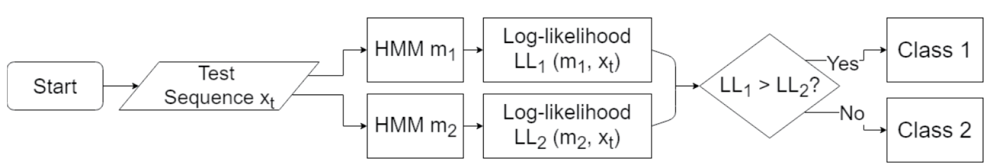
Two variables from the NEST dataset, road surface condition (good or poor conditions) and road alignment (straight or curved road), were combined to create the "motor control difficulty" categories: lower and higher. Baselines with good surface conditions and straight roads were categorized as having lower motor control difficulty, the rest were categorized as having higher motor control difficulty. Based on these categorizations and after removing baselines with unknown/missing data, 617 baselines were labeled as having "lower motor control difficulty" and 315 baselines were labeled as having "higher motor control difficulty." Overall, the total of 932 baselines corresponded to a total of 186,998 frames of vehicle-based data recorded at 10 Hz.

Hidden Markov Models (HMMs)

After the removal of baselines that had missing data as reported above, data had to be further processed for modeling purposes. Because the GPS speed was sampled at a lower rate (1 Hz) than the other three predictor variables (10 Hz), GPS speed data was interpolated/extrapolated over the 20 second baseline interval to fill in missing entries and to bring it up to the same frequency as the other three predictor variables. The same technique was used when the other three predictor variables had missing values for a baseline (120 missing frames were generated for steering wheel position, and 3047 both for x- and y-acceleration). If a baseline had no values or a single value recorded for a predictor, interpolation/extrapolation was not possible, and that baseline was removed from the analysis. After this data processing step, the number of baselines that could be used in HMMs was further reduced as reported in the last column of Table 3. Significant reductions occurred in the baseline data used for distraction classification, given that the predictors GPS speed and steering wheel position had high amounts of missing data.

HMMs were developed for each classification problem in Python 3.2 using the hmmlearn library. The classifications were conducted with generative models using first order Gaussian HMM. This approach creates two HMMs (m1 and m2 in Figure 1) – one for each classification level (e.g. one for secondary task engagement and one for no secondary task engagement) – but does not need to explicitly specify the HMM model parameters. The Baum-Welch algorithm (Baum, Petrie, Soules, & Weiss, 1970), an expectation maximization algorithm that uses Maximum Likelihood Estimation (MLE) (Dempster, Laird, & Rubin, 1977), was used to construct the models. The baselines were randomly split into 80%-20% training-test sets. All three problems had unbalanced classes, a situation known to hinder the performance of learning classifier systems (Sun, Wong, & Kamel, 2009). To balance the training sets, the classes that had more baselines were randomly undersampled using the RandomUnderSampler method from the imbalanced-learn library in Python. After the models were trained, test sequences (x_t) were fed into the HMMs as shown in Figure 1. Log-likelihoods were computed for each model, and the test sequence was assigned to the model that gave the higher log-likelihood value.

Figure 1. The HMM classification process

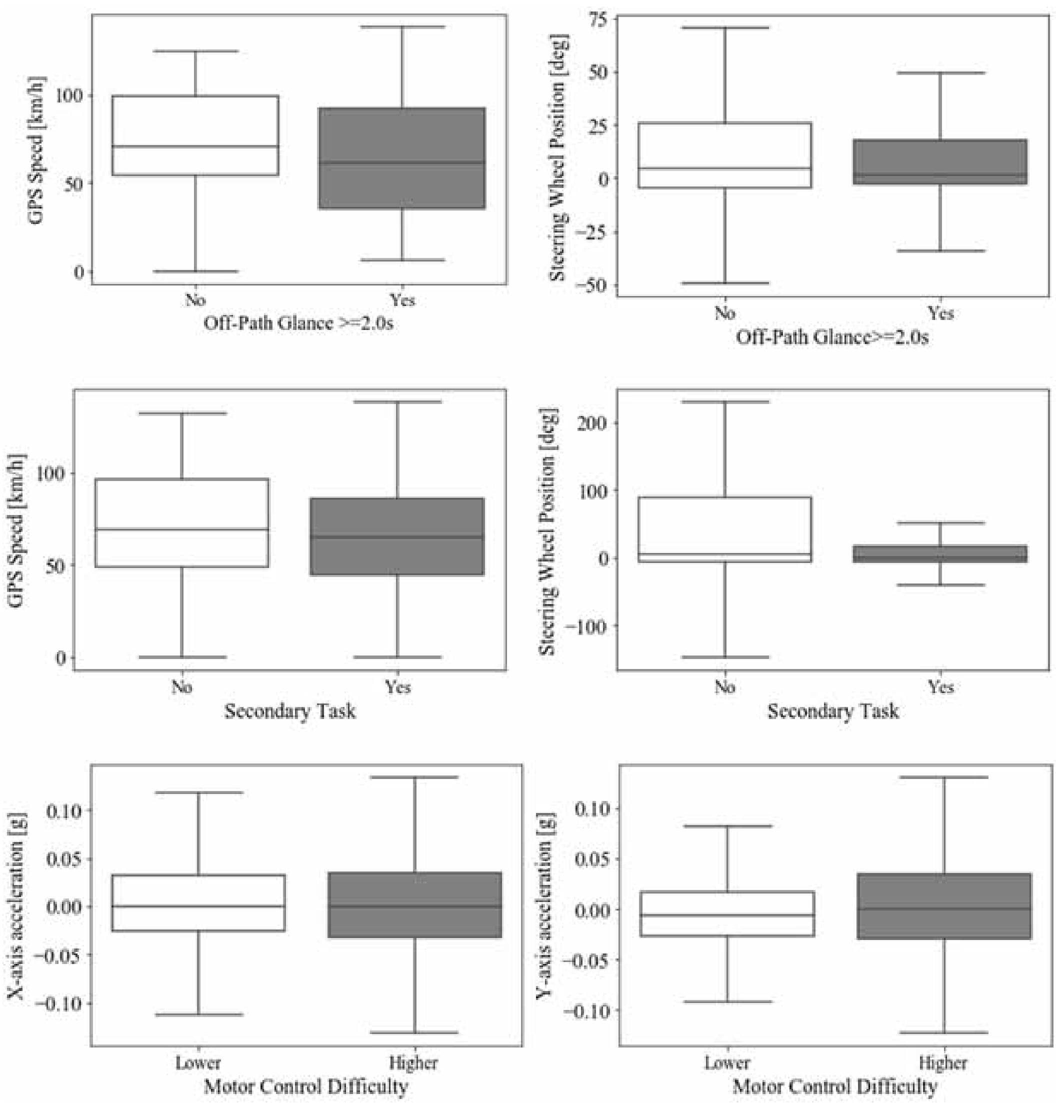


RESULTS

Data Properties

Figure 2 presents boxplots of the predictor variables from a randomly selected training dataset that was used in modeling. Overall, visual inspections of these plots along with normal quantile-quantile plots of the data that are not presented here suggest approximately normal distributions for the data. In particular, the variability of the steering wheel position appears to decrease under distracted states; the GPS speed also tends to decrease. These trends are in line with previous research that reported a reduction in speed (Engström, Johansson, & Östlund, 2005) and variability of steering input (Kircher & Ahlstrom, 2010) under distraction. As for motor control difficulty, the variability of both lateral and longitudinal acceleration appears to be higher during higher motor control difficulty situations.

Figure 2. Boxplots of predictor variables across different classes; boxplots depict the five-number summary



Classification

Each classification problem was repeated 10 times with randomly split training and test datasets in order to examine the robustness of the models in uncovering data patterns. The classification performance of the HMMs is presented in Table 4. The number of hidden states were determined empirically; 3 states were chosen for long off-path glance classification, whereas 2 states were chosen for the other two classification problems.

The accuracy of the models to classify distraction (i.e. presence of long off-path glances or secondary task engagement) was relatively high at 77%. However, these models were performing at the chance level or worse in classifying non-distraction cases. As for motor control difficulty, the accuracy levels were 60% and 67% for classifying higher and lower difficulty cases, respectively.

DISCUSSION

Hidden Markov models (HMMs) were built based on a naturalistic driving dataset, to detect instances of driver distraction as well as high levels of environmental demand from vehicle-based data. In particular, the Naturalistic Engagement in Secondary Tasks (NEST) dataset, a subset of SHRP2 data, was utilized to classify whether drivers had long off-path glances (2 seconds or longer), whether they were engaged in a secondary task, and whether the motor control difficulty associated with road conditions and alignment was relatively high. Specifically, GPS speed and steering wheel position were used to classify the two distraction indicators, whereas lateral and longitudinal acceleration were used to classify motor control difficulty. Overall, the models performed relatively well at around 77% accuracy for detecting existence of long off-path glances and secondary task engagement. However, they failed to achieve a high accuracy rate for detecting cases of no long off-path glances and no secondary task engagement. As for motor control difficulty, the models achieved 60 and 67% accuracy rates for higher and lower motor control difficulty levels, performing better than chance in labelling both classes.

It appears that although driving behavior becomes more specific (i.e. lower speed and significantly reduced variability of the steering wheel angle) when drivers have long off-path glances or when they are engaging in secondary tasks, these specific driving behaviors are also observed when drivers are not engaged in secondary tasks or are not looking away from the road for extended periods of time. Due to this overlap, the models may be failing to perform well for accurately classifying the “no” secondary task and the “no” long off-path glance situations, which may translate to false alarms in driver state detection systems. Therefore, although vehicle-based measures can be informative, they may need to be combined with other signals in order to boost their performance. Another potential

Table 4. Classification accuracy (average and standard deviation across 10 repetitions) of HMMs for the three classification problems

Predicted Variable	Predictors	Classes	Accuracy	
			Average	Standard Deviation
Long Off-path Glances ($\geq$ 2 sec)	GPS speed and steering wheel position	Yes	77.14%	13.80%
		No	51.38%	9.54%
Secondary Task Engagement	GPS speed and steering wheel position	Yes	77.04%	13.73%
		No	42%	16.87%
Motor Control Difficulty	x- and y-acceleration	Higher	59.82%	7.74%
		Lower	67.22%	4.48%

reason for low accuracies for the "no" classes might be due to the limited size of the data we could utilize in our models. A significant amount of the baseline cases was lost due to missing values in steering wheel position and GPS speed. Given that the classification accuracy depends on the richness of the training dataset, using a larger dataset such as the entire SHRP2 data or the inclusion of more variables such as headway distance and vehicle lane position that describe drivers' behavior more comprehensively may also result in significant improvements in accuracy. Performance boost may also be achieved by using the Synthetic Minority Over-Sampling Technique (SMOTE) (Chawla, Bowyer, Hall, & Kegelmeyer, 2002) to generate new data samples from the minority class and create balanced datasets.

Classification accuracy may also be improved by utilizing other algorithmic approaches. Other models suitable for capturing temporal dynamics, such as hidden semi-Markov models (Johnson & Willsky, 2012), recurrent neural networks (Schmidhuber, 2015), and Bayesian network models, should be explored in future research. It should also be noted that our models were not trained on individual drivers' data. Therefore, as expected, our accuracy levels are similar to or lower than other studies which trained classifiers on driver specific data (Kircher & Ahlstrom, 2010; Li et al., 2018). If driver state detection systems can be trained on the individual driver, then they should perform better for that specific driver.

It should also be noted that although we have tried to capture two different indicators of driver distraction that are known to affect crash risk and driving performance, these indicators may not fully capture all distracted driving states. For example, drivers may be lost in thought and their reaction times may be delayed and these states may also be important to capture through a driver state detection system. Further, it was expected that lateral and longitudinal acceleration would be indicative of motor control difficulty associated with the road environment. This expectation was confirmed given the relatively good performance of the models for classifying motor control difficulty based on acceleration data. In general, better system awareness of the driving environment can lead to more effective and intelligent distraction mitigation strategies and further research is needed to characterize environmental demand beyond motor control difficulty.

CONCLUSION

In conclusion, using hidden Markov models built on naturalistic driving data, vehicle-based measures (speed and steering wheel position) were found to predict indicators of driver distraction (secondary task engagement and long off-path glances) with an accuracy of 77%. A hidden Markov model utilizing lateral and longitudinal acceleration to predict higher motor control difficulty was found to have an accuracy of 60%. The performance of such models can be improved by including additional vehicle-based measures as predictors, using different techniques for creating balanced classes in the training datasets, or utilizing different modeling algorithms. Overall, our results suggest that vehicle-based measures are promising for use in the prediction of driver distraction and environmental demand. Such predictions can be used in driver assistance systems that support drivers during periods of distraction and high environmental demand.

ACKNOWLEDGMENT

This work was sponsored by the Toyota Collaborative Safety Research Center (CSRC). We would like to thank Joshua Domeyer, Heishiro Toyoda, James Foley, and Chuck Gulash from Toyota CSRC for their valuable feedback. We gratefully acknowledge the help provided by VTTI (Julie Cook, Suzie Lee, Justin Owens, and Miguel Perez) regarding our inquiries about the data. Earlier modeling attempts on this dataset have been reported in the Proceedings of the 10th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications.

REFERENCES

Achten, J., & Jeukendrup, A. E. (2003). Heart rate monitoring: Applications and limitations. *Sports Medicine*, *33*(7), 517–538. doi:10.2165/00007256-200333070-00004 PMID:12762827

Aghaei, A. S., Donmez, B., Liu, C. C., He, D., Liu, G., Plataniotis, K. N., & Sojoudi, Z. et al. (2016). Smart driver monitoring: When signal processing meets human factors. *IEEE Signal Processing Magazine*, *33*(6), 35–48. doi:10.1109/MSP.2016.2602379

Atiquzzaman, M., Qi, Y., & Fries, R. (2018). Real-time detection of drivers' texting and eating behavior based on vehicle dynamics. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 594–604. doi:10.1016/j.trf.2018.06.027

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, *41*(1), 164–171. doi:10.1214/aoms/1177697196

Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident; Analysis and Prevention*, *40*(4), 1282–1293. doi:10.1016/j.aap.2008.01.009 PMID:18606257

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. doi:10.1613/jair.953

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the *EM* algorithm. *Journal of the Royal Statistical Society. Series B. Methodological*, *39*(1), 1–22. doi:10.1111/j.2517-6161.1977.tb01600.x

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 of the United States of America*, *113*(10), 2636–2641. doi:10.1073/pnas.1513271113 PMID:26903657

Dingus, T. A., Hankey, J. M., Antin, J. F., Lee, S. E., Eichelberger, L., & Stulce, K. … Loren, M. (2014). Naturalistic Driving Study: Technical Coordination and Quality Control. Washington, D.C.: Transportation Research Board.

Donmez, B., Boyle, L. N., & Lee, J. D. (2007). Safety implications of providing real-time feedback to distracted drivers. *Accident; Analysis and Prevention*, *39*(3), 581–590. doi:10.1016/j.aap.2006.10.003 PMID:17109807

Donmez, B., Boyle, L. N., & Lee, J. D. (2008). Mitigating driver distraction with retrospective and concurrent feedback. *Accident; Analysis and Prevention*, *40*(2), 776–786. doi:10.1016/j.aap.2007.09.023 PMID:18329433

Donmez, B., & Liu, Z. (2015). Associations of distraction involvement and age with driver injury severities. *Journal of Safety Research*, *52*, 23–28. doi:10.1016/j.jsr.2014.12.001 PMID:25662879

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*(2), 97–120. doi:10.1016/j.trf.2005.04.012

Fridman, L., Toyoda, H., Seaman, S., Seppelt, B., Angell, L., & Lee, J., … Reimer, B. (2016). What can be predicted from six seconds of driver glances? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 2805–2813).

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors*, *48*(1), 196–205. doi:10.1518/001872006776412135 PMID:16696268

International Organization for Standardization. (2007). *Road vehicles -- Ergonomic aspects of transport information and control systems -- Occlusion method to assess visual demand due to the use of in-vehicle systems* (No. 16673)

Jain, A., Koppula, H. S., Raghavan, B., Soh, S., & Saxena, A. (2015). Car that knows before you do: Anticipating maneuvers via learning temporal driving models. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3182–3190). doi:10.1109/ICCV.2015.364

Johnson, M. J., & Willsky, A. S. (2012). Bayesian nonparametric hidden semi-Markov models. *Journal of Machine Learning Research*, *14*, 673–701.

Kircher, K., & Ahlstrom, C. (2010). Predicting visual distraction using driving performance data. *Annals of Advances in Automotive Medicine*, *54*, 334–342. PMID:21050615

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data*. Washington, D.C.: National Highway Traffic Safety Administration.

Lee, J., Muñoz, M., Fridman, L., Victor, T., Reimer, B., & Mehler, B. (2018). Investigating the correspondence between driver head position and glance location. *PeerJ Computer Science*, *4*, e146. doi:10.7717/peerj-cs.146

Li, Z., Bao, S., Kolmanovsky, I. V., & Yin, X. (2018). Visual-manual distraction detection using driving performance indicators with naturalistic driving data. *IEEE Transactions on Intelligent Transportation Systems*, *19*(8), 2528–2535. doi:10.1109/TITS.2017.2754467

Martin, S., Vora, S., Yuen, K., & Trivedi, M. M. (2018). Dynamics of driver's gaze: explorations in behavior modeling & maneuver prediction. arXiv:1802.00066, 8858(c)

McCall, J. C., & Trivedi, M. M. (2004). Visual context capture and analysis for driver attention monitoring. In *Proceedings of the IEEE Intelligent Transportation Systems Conference* (pp. 332–337). doi:10.1109/ITSC.2004.1398920

National Highway Traffic Safety Administration. (2018). *Distracted Driving 2016 (No. DOT HS 812 517)*. Washington, D.C.: National Highway Traffic Safety Administration.

Norrman, J., Eriksson, M., & Lindqvist, S. (2000). Relationships between road slipperiness, traffic accident risk and winter road maintenance activity. *Climate Research*, *15*, 185–193. doi:10.3354/cr015185

Owens, J. M., Angell, L., Hankey, J. M., Foley, J., & Ebe, K. (2015). Creation of the Naturalistic Engagement in Secondary Tasks (NEST) distracted driving dataset. *Journal of Safety Research*, *54*, 33.e29–e36. doi:10.1016/j.jsr.2015.07.001 PMID:26403898

Pentland, A., & Liu, A. (1999). Modeling and prediction of human behavior. *Neural Computation*, *11*(1), 229–242. doi:10.1162/089976699300016890 PMID:9950731

Regan, M. A., Lee, J. D., & Young, K. (2008). *Driver Distraction: Theory, Effects, and Mitigation*. Boca Raton: CRC Press. doi:10.1201/9781420007497

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. doi:10.1016/j.neunet.2014.09.003 PMID:25462637

Stanton, N. A., & Young, M. S. (2005). Driver behaviour with adaptive cruise control. *Ergonomics*, *48*(10), 1294–1313. doi:10.1080/00140130500252990 PMID:16253946

Sun, Y., Wong, A. K. C., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, *23*(4), 687–719. doi:10.1142/S0218001409007326

Taelman, J., Vandeput, S., Spaepen, A., & Van Huffel, S. (2009). Influence of mental stress on heart rate and heart rate variability. In J. Vander Sloten, P. Verdonck, M. Nyssen, & J. Haueisen (Eds.), *4th European Conference of the International Federation for Medical and Biological Engineering* (Vol. 22, pp. 1366–1369). Springer doi:10.1007/978-3-540-89208-3_324

Tango, F., & Botta, M. (2013). Real-time detection system of driver distraction using machine learning. *IEEE Transactions on Intelligent Transportation Systems*, *14*(2), 894–905. doi:10.1109/TITS.2013.2247760

Tchankue, P., Wesson, J., & Vogts, D. (2011). The impact of an adaptive user interface on reducing driver distraction. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '11,* Salzburg, Austria (p. 87). ACM Press. doi:10.1145/2381416.2381430

Yamato, J., Ohya, J., & Ishii, K. (1992). Recognizing human action in time-sequential images using hidden Markov model. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* Champaign, IL (pp. 379–385). IEEE Comput. Soc. doi:10.1109/CVPR.1992.223161

Yang, G., Lin, Y., & Bhattacharya, P. (2010). A driver fatigue recognition model based on information fusion and dynamic Bayesian network. *Information Sciences*, *180*(10), 1942–1954. doi:10.1016/j.ins.2010.01.011

Young, K. L., & Lenné, M. G. (2010). Driver engagement in distracting activities and the strategies used to minimise risk. *Safety Science*, *48*(3), 326–332. doi:10.1016/j.ssci.2009.10.008

Dina Kanaan is a PhD student at the Human Factors and Applied Statistics Laboratory in the Department of Mechanical and Industrial Engineering at the University of Toronto. Dina's research interests include human factors in driving applications, driver attention and distraction, automated driving, driver assistance systems, and driver-automation interaction. Dina obtained her Bachelor of Engineering Degree in Electrical and Computer Engineering from the American University of Beirut (AUB) in 2012 and her Master of Engineering Management from AUB in 2018. Prior to joining the University of Toronto, Dina worked as a software developer in Beirut, Lebanon for five years.

Suzan Ayas is a MASc student at the University of Toronto Mechanical and Industrial Engineering Department.

Birsen Donmez received her B.S. degree in mechanical engineering from Bogazici University, Istanbul, Turkey, in 2001, and M.S. and Ph.D. degrees in industrial engineering in 2004 and 2007, respectively, and M.S. degree in statistics in 2007 from the University of Iowa, Iowa City, IA, USA. She is currently an Associate Professor at the University of Toronto, Mechanical & Industrial Engineering and is the Canada Research Chair in Human Factors and Transportation. Donmez serves on multiple Transportation Research Board committees and as an associate editor for IEEE Transactions on Human-Machine Systems.

Martina Risteska obtained her Master of Applied Science (MASc) degree in Industrial Engineering from University of Toronto in November 2018. Her research interests are centered around data analytics in the field of autonomous driving, probabilistic modeling of human behaviour, and cognitive robotics. In addition to her MASc degree, Martina holds a Bachelor of Science degree in Systems Engineering and Automation from the Faculty of Electrical Engineering and Information Technologies in Skopje, Macedonia.

Joyita Chakraborty is a Master of Applied Science candidate at University of Toronto Mechanical and Industrial Engineering Department.