Accident Analysis and Prevention 123 (2019) 274–281

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

A hierarchical machine learning classification approach for secondary task identification from observed driving behavior data

Osama A. Osman[a,*,1,2], Mustafa Hajij[b,1,3], Sogand Karbalaieali[a], Sherif Ishak[c,4]

[a] Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA 70803, United States
[b] Department of Computer Science and Engineering, University of South Florida, Tampa, FL 33620, United States
[c] Department of Civil and Environmental Engineering, University of Alabama, Huntsville, AL 35899, United States

ARTICLE INFO

Keywords:
Distracted driving
Secondary tasks
Detection
Identification
Driving behavior
Ensemble tree
Machine learning
Accident investigations
In-vehicle systems

ABSTRACT

According to NHTSA, more than 3477 people (including 551 non-occupants) were killed and 391,000 were injured due to distraction-related crashes in 2015. The distracted driving epidemic has long been under research to identify its impact on driving behavior. There have been a few attempts to detect drivers' engagement in secondary tasks from observed driving behavior. Yet, to the authors' knowledge, not much effort has been directed to identify the types of secondary tasks from driving behavior parameters. This study proposes a bi-level hierarchical classification methodology using machine learning to identify the different types of secondary tasks drivers are engaged in using their driving behavior parameters. At the first level, drivers' engagement in secondary tasks is detected, while at the second level, the distinct types of secondary tasks are identified. Comparative evaluation is performed between nine ensemble tree classification methods to identify three types of secondary tasks (hand-held cellphone calling, cellphone texting, and interaction with an adjacent passenger). The inputs to the models are five driving behavior parameters (speed, longitudinal acceleration, lateral acceleration, pedal position, and yaw rate) along with their standard deviations. The results showed that the overall secondary task detection accuracy ranged from 66% to 96%, except for the Decision Tree that was able to detect engagement in secondary tasks with a high accuracy of 99.8%. For the identification of secondary tasks types, the overall accuracy ranged from 55% to 79%, with the highest accuracy of 82.2% achieved by the Random Forest method. The findings of the paper show the proposed methodology promising to (1) characterize drivers' engagement in unlawful secondary tasks (such as texting) as a counter measure to prevent crashes, and (2) alert drivers to pay attention back to the main driving task when risky changes to their driving behavior take place.

1. Introduction

Traffic safety has always been a major concern for government officials, transportation researchers, and roadway users. Distracted driving behavior is one of the main issues that compromise safety. The NHTSA crash data indicate that in 2015, at least 3477 people (including 551 non-occupants) were killed and 391,000 were injured in crashes that were deemed to involve distraction (Anon., 2014). Causes of distracted driving involve several activities such as eating, cellphone use, and talking to passengers among others. Studies have shown that cellphone use (calling or texting) is amongst the high risk activities, as it is highly likely to increase crash risk by 4 times (Hosking et al., 2005; McEvoy et al., 2005). Additionally, studies have shown that engaging in conversations while driving, whether on the phone or with passengers in the car, is a high risk activity that significantly increases the likelihood of crashes (Consiglio et al., 2003; Horrey and Wickens, 2004).

Driving simulator studies and naturalistic driving studies are two ways that have been used to investigate the impact of distracted driving. Experiments in driving simulators are easier to control and the associated data collection process is relatively easier and non-invasive. Nevertheless, the controlled settings and simulation environments provide a lesser degree of realism compared to naturalistic driving

* Corresponding author.
E-mail addresses: oosman@vtti.vt.edu (O.A. Osman), hajij.1@osu.edu (M. Hajij), skarba1@lsu.edu (S. Karbalaieali), sishak@odu.edu (S. Ishak).
[1] Contributed equally to this work,
[2] Center for Sustainable Mobility, Virginia Tech Transportation Institute, Virginia Polytechnic Institute and State University, 3500 Transportation Research Plaza, Blacksburg, VA, 24060, 540-231-0278, United States.
[3] Department of Computer Science and Engineering, Ohio State University, 2015 Neil Ave, Columbus, OH 43210, United States.
[4] Department of Civil and Environmental Engineering, Old Dominion University, 135 Kaufman Hall, Norflok, VA, 23529, United States.

https://doi.org/10.1016/j.aap.2018.12.005
Received 2 May 2018; Received in revised form 9 November 2018; Accepted 6 December 2018
Available online 13 December 2018

studies (NDS). NDS data include observations of drivers in their own vehicles while performing their regular commutes. Among the first conducted studies over the past few years using naturalistic driving data are two large-scale studies by Stutts et al. (2005) and (2003) to measure causes and consequences of distraction. The 100-Car study is one of the largest scale NDS which included about 3 million vehicle miles and yielded about 42,300 data hours, 82 crashes, 761 near-crashes, and 8295 critical incidents (Dingus et al., 2006). The 100-Car study data has been used in several behavioral-based studies such as validating the use of near-crashes as crash surrogates (Jonasson and Rootzén, 2014), evaluating safety critical braking events (Bagdadi, 2013), predicting high-risk drivers based on their demographic, personality, and driving characteristic data (Guo and Fang, 2013), modeling of car-following behavior (Sangster et al., 2013), and examining the effect of driver inattention (Wong and Huang, 2013).

The SHRP 2 NDS is the largest scale naturalistic driving database available to date. The SHRP2 NDS database provide ample opportunity for a better understanding of driving behavior and impact of several factors (including distraction) on driving safety (Anon., 2013). The millions of video hours collected in the study allowed researchers to quantify the impact of distractions on drivers' head position and eye and mouth states (Smith et al., 2017). The driving behavior and driving state data collected in the SHRP2 NDS helped researchers understand and quantify the effect of secondary tasks and driver behavior on crash and near-crash risk (Ashley et al., 2017; Ye et al., 2017a). The driving state data also allowed researchers to better understand heterogeneity of crashes (Knipling, 2017). The SHRP2 NDS trip data was used for driving cycle development and estimation of fuel consumption and vehicle emissions (Sun et al., 2017). Finally, the SHRP2 NDS data gave researchers the opportunity to understand the impact of roadway lighting on driver behavior (Li et al., 2017a), heavy rain on drivers' lane keeping ability (Ghasemzadeh and Ahmed, 2017), adverse weather conditions on speed and headway behavior (Ahmed and Ghasemzadeh, 2017), and roadway improvement projects on safety (Li et al., 2017b).

Naturalistic driving data have also been used by many researchers to investigate the impact of distracted driving on safety. For instance, Klauer et al. (2006) found that drivers' engagement in secondary tasks increase crash risk by two times compared to normal driving. In another study, Victor et al. (2015) explored the crash risk associated with drivers' engagement in secondary tasks. The study found that most secondary tasks significantly increase the likelihood of crashes, with texting identified as the secondary task with the highest crash risk. Dingus et al. (2016) evaluated the likelihood of drivers' engagement in secondary tasks and the associated crash risk. In their study, Dingus et al. evaluated the impact of several secondary tasks including interaction with a passenger, talking on a handheld cell phone, eating, and adjusting the radio, among others. The study concluded that drivers tend to be engaged in at least one secondary task during 52% of the time while driving, which increases crash risk by at least two times.

Studies with the naturalistic driving data confirm the long proven fact that distracted driving is a risky behavior that significantly increases crash risk. While most of the literature focused on investigating the impact of distracted driving and secondary task engagement on driver behavior and crash risk, minimal effort was directed to investigate whether distracted driving could be detected from observed driving behavior data. For instance, researchers applied Multiple Logistic Regression (MLR) to detect drivers' engagement in secondary tasks using information about observed driving behavior (Jenkins et al., 2017). The study showed that traditional statistical techniques are not efficient for this type of modeling. The main challenge stemmed from the fact that relationships involving driver behavior tend to be more complex and highly nonlinear.

Machine learning techniques have long proven promising and more effective in solving several complex and nonlinear problems (Bakhit et al., 2017; Hofleitner et al., 2012; Moazenzadeh et al., 2018; Mousa et al., 2018; Osman et al., 2018; Vanajakshi and Rilett, 2007; Weiss and Kulikowski, 1991; Zhang and Haghani, 2015). Given this promising performance of machine learning techniques, Ye et al. (2017b) investigated the possibility of detecting drivers' engagement in secondary tasks through the application of artificial neural networks. The study found that neural networks' modeling can with high accuracy help detect instances of drivers performing phone calls, texting, and conversation with a passenger. The study, however, did not investigate whether the secondary tasks may affect driver behavior differently depending on their types. This is rather important to differentiate between unlawful secondary tasks and permitted yet distracting tasks in accident investigations. Thus, this study is an attempt to answer two questions: (a) "does driving behavior change in a different way depending on the type of secondary task drivers perform?", and if the answer is yes, (b) can machine learning techniques be used to detect the types of the secondary tasks drivers are engaged in from their observed driving behavior?" To answer these questions, several machine learning classifiers are applied on SHRP2 NDS driving behavior parameters for identification of secondary task types. The current study investigates the possibility of answering these questions when knowledge of only driving behavior parameters of the subject vehicle (the vehicle in which the secondary task engagement takes place) is available, regardless of the vehicle type or the roadway geometric characteristics. Studying the effect of roadway and geometric characteristics is not within the scope of this study.

The remainder of the paper presents a description of the data used in the study followed by the study methodology. A discussion of the models used in the study and modeling results is then presented, and finally a detailed discussion of the study results and conclusions is provided.

2. Data description and processing

The study hypothesis is that different types of secondary tasks affect driving behavior differently and hence can be identified from observed driving behavior data. To test this hypothesis, data about types of secondary tasks and the associated driving behavior parameters are extracted from the SHRP2 NDS database. Only hand-held cellphone calling and texting as well as conversation with adjacent passengers are the secondary tasks considered in the study since they are identified among the highest crash risk tasks by several studies (Hosking et al., 2005; McEvoy et al., 2005; Consiglio et al., 2003; Horrey and Wickens, 2004). The time-series driving speed, lateral acceleration, longitudinal acceleration, yaw rate, and pedal position are extracted from the SHRP2 NDS database and used as measures of the driving behavior associated with each of the selected secondary tasks. The five parameters are extracted at a resolution of 0.1 s for a sample of 100 baseline (no crash or near crash) events for each type of secondary task. Additionally, the five parameters are extracted for a total of 100 baseline events of normal driving (no secondary tasks). These five parameters are identified by several studies to be most influenced when drivers are distracted (Klauer et al., 2006; Jenkins et al., 2017). Table 1 shows the units and definitions of each of the five parameters, and Table 2 shows their summary statistics grouped by the secondary task type.

As shown in Table 2, the total number of observations is 26,028, 25,433, 25,644, and 61,071 observations for the hand-held calling, texting, interaction with adjacent passengers, and normal driving categories, respectively. These numbers are based on the 0.1-second data resolution collected over periods of 20–26 s for calling, texting, and interaction with adjacent passengers, and 60–61 s for normal driving. These observations were collected from a total of 373 drivers, divided as 95 drivers calling, 96 drivers texting, 84 drivers engaging in conversation with the adjacent passenger, and 98 drivers not engaged in any secondary task. Some drivers were engaged in multiple secondary task or normal driving events. Looking at the values in Table 2, it is clear that there are some differences between the summary statistics associated with each category. While these differences seem marginal,

Table 1
Driving Behavior Parameters.

Parameter	Units	Description
Driving Speed	km/hr	Vehicle speed indicated on speedometer collected from network.
Lateral Acc. (accel_y)	g	Vehicle acceleration in the lateral direction versus time.
Longitudinal Acc. (accel_x)	g	Vehicle acceleration in the longitudinal direction versus time.
Yaw Rate (gyro_z)	deg/sec	Vehicle angular velocity around the vertical axis.
Pedal Position	%	Position of the accelerator pedal

which might indicate that the driving patterns associated with the different categories could be similar, further investigation using machine learning algorithms is required.

While sampling the 400 events (100 events for each secondary task and 100 events for normal driving), the data are cleaned from any deficient events that have zero observations throughout the entire event duration. Missing observations are then interpolated based on the following rules: (1) the previous and the following observed speeds to a missing observation(s) should be greater than 5 km/hr (idle speed), (2) the previous and the following observed longitudinal and lateral accelerations should be greater than 0.1 m/s^2, (3) the preceding and the following yaw rates should be higher than 3 °/sec, and (4) the pedal position should be higher than 0.5%. These rules are determined based on a thorough investigation of the data and the available values in the NDS database. The thresholds used in these rules were determined based on the pattern of values in the data. For instance, when looking at the data, it was very frequent to observe zero values after acceleration values of 0.1 m/s^2, which could be an indicator that the observed zero is an actual zero acceleration value rather than being a missing value. When any rule is not satisfied for the parameter being interpolated, missing observations are treated as observed actual zeros.

Since changes in driving behavior take place over time, measures of driving behavior should be studied as a pattern rather than independent observations. To do so, a moving time window is applied along the time-series observations of each parameter. The standard deviation is calculated for each parameter over each time window of a one-second length (10 time steps of 0.1 s each) to capture minimal changes in the driving behavior patterns. As shown in Fig. 1, the standard deviation is calculated for the observations falling within each one-second window and placed in a new dataset of standard deviations. The window is then moved one time-step (0.1 s) at a time over the entire event data to calculate the following standard deviation values. This is repeated for each parameter until the window moves over all observations in all events.

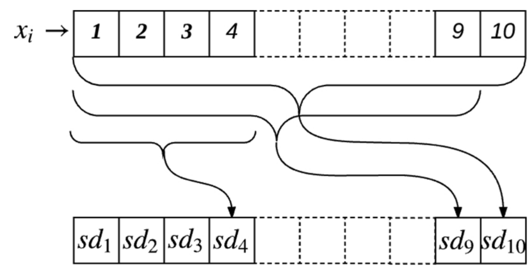

Fig. 1. Moving window for standard deviation calculation.

To apply the different machine learning algorithms, the data is first standardized, using Eq. (1), and normalized, using Eq. (2), such that all

Table 2
Summary Statistics of the Driving Parameters.

Cellphone Calling (# of observations = 26,028)

Statistical Measure	Speed	Lateral Acceleration	Longitudinal Acceleration	Throttle position	Yaw Rate
Mean	52.840	−0.004	−0.054	11.288	−0.181
Standard Deviation	36.912	0.057	0.491	11.904	3.738
Maximum	134.880	0.481	0.377	58.000	31.219
Minimum	0.000	−0.481	−5.092	0.000	−34.146

Cellphone Texting (# of observations = 25,433)

Statistical Measure	Speed	Lateral Acceleration	Longitudinal Acceleration	Throttle position	Yaw Rate
Mean	53.670	−0.003	−0.004	11.930	−0.019
Standard Deviation	36.953	0.057	0.065	12.233	2.863
Maximum	132.730	0.537	0.348	86.270	88.453
Minimum	0.000	−0.563	−0.508	0.000	−28.617

Interaction (# of observations = 25,644)

Statistical Measure	Speed	Lateral Acceleration	Longitudinal Acceleration	Throttle position	Yaw Rate
Mean	55.400	−0.005	−0.053	12.200	−0.121
Standard Deviation	36.358	0.060	0.496	12.156	3.508
Maximum	130.500	0.516	0.415	64.000	48.129
Minimum	0.000	−0.352	−5.037	0.000	−36.097

Normal Driving(# of observations = 61,071)

Statistical Measure	Speed	Lateral Acceleration	Longitudinal Acceleration	Throttle position	Yaw Rate
Mean	53.910	0.004	−0.003	11.040	0.183
Standard Deviation	39.223	0.069	0.066	11.630	3.726
Maximum	126.330	0.687	0.418	79.220	31.544
Minimum	0.000	−0.597	−0.412	0.000	−33.170

features will have comparable ranges of values. This is an important step in machine learning to avoid domination of any of the features over the others when training a machine learning algorithm.

$$x_{new} = \frac{x - \mu}{\sigma} \tag{1}$$

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

3. Methodology and analysis

Secondary task identification can be treated as a supervised machine learning classification problem. A supervised machine learning classifier takes a data set that consists of training examples (x_1, y_1) to (x_N, y_N) as an input, where $\{x_i\}_{i=1}^{N} \subseteq R^P$ is a set of feature vectors in the feature space R^P and y_i is a label that takes value in a finite set Y. A classifier then divides the feature space along hypersurfaces, also called decision boundaries, into regions that best separate the data points $x_1, \ldots x_N$ according to their labels.

In this study, the driving behavior data are labeled to represent four distinct classes: hand-held cellphone calling, cellphone texting, interaction with passengers, and normal driving. A total of 10 features are used (speed, speed standard deviations, lateral acceleration, lateral acceleration standard deviations, longitudinal acceleration, longitudinal acceleration standard deviations, yaw rate, yaw rate standard deviations, pedal position, and pedal position standard deviations) as inputs to the different classifiers. As mentioned earlier, driving behavior is studied as a pattern over time rather than individual observations. However, one of the main study hypothesis is that the reason for a specific driving behavior pattern taking place at certain values of driving behavior performance variables (e.g. speed = 60mph) is different from when that same pattern takes place at different values of the performance variables (e.g. speed = 35mph). Therefore, the standard deviations dataset is combined with the five driving behavior parameters in this study. Using the Python Sicket-learn package, several classification methods are tested and compared including K Nearest Neighbor (KNN), Random Forest, Support Vector Machine (SVM), Decision Trees, Gaussian Neighborhood, Adaptive Boost (AdaBoost), Multilayer Perceptron (MLP), and Quadratic Discrimination Analysis (QDA). These ensemble tree classifiers were widely implemented in several transportation applications such as traffic accident severity, driver injury severity (Chang and Chien, 2013), characterization of precrash maneuvers (Harb et al., 2009), analysis of gap acceptance behavior (Nagalla et al., 2017), and estimating work zone capacity (Weng and Meng, 2012). To develop the models, the data are randomly divided into two sets, 80% for training and 20% for testing, and the 5-fold cross validation is applied.

Initial results show that almost all models give relatively poor detection accuracy when trying to classify the data into the four different classes simultaneously. Table 3 illustrates that the highest classification accuracy can be obtained with the Decision Tree classifier. While this accuracy may not be very poor, it is not considered reliable for detecting engagement in secondary tasks that can compromise driver safety.

The reason for the poor classification accuracy can be explained by investigating the geometry of the dataset at hand. To illustrate what is meant by data geometry, the dataset is projected nonlinearly to the 2-D plane using one of the dimensionality reduction techniques. In this study, the Principal Component Analysis (PCA) is applied. In PCA, the original coordinates of the features are changed and the axes of the new coordinate system are chosen to match the directions of the highest variance in the data. In a planner projection the data are projected on the axes with the highest two variance directions. The results in Fig. 2(a) indicate that the normal driving class (labeled with red color) is clustered together, while the classes associated with engagement in the three types of secondary tasks overlap with each other around and minimally with the normal driving class. This overlapping indicates a difficulty in trying to build a decision boundary that separates the four classes. Indeed, poor accuracy results are obtained for all classifiers in Table 1.

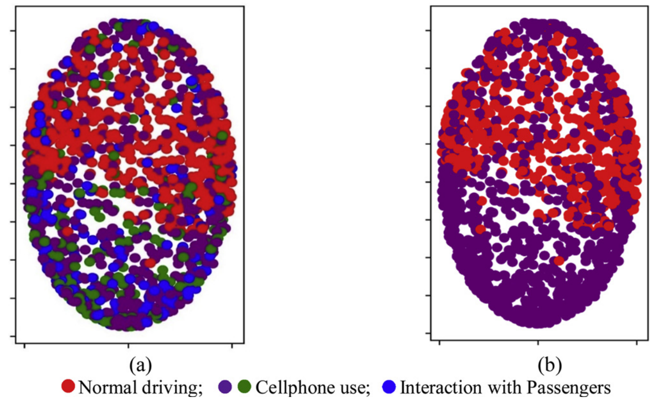

Fig. 2. 2-D PCA projection of data; (a) four Classes, (b) two classes.

The fact that the normal driving class can be distinct in Fig. 2(a) led to considering binary classification, where one part of the data represents normal driving and the other represents engagement in all types of secondary tasks. As depicted in Fig. 2(b), it seems more plausible to differentiate between normal driving behavior and the behavior associated with engagement in secondary tasks. Therefore, a bi-level hierarchical classification approach is adopted to detect engagement in each type of secondary task. More specifically, using supervised classification techniques, engagement in secondary tasks is detected considering only two classes (normal driving and secondary task engagement) at the first level. Once engagement in secondary tasks is detected, supervised classification techniques are used to identify the type of the secondary task at the second level. Fig. 3 shows the proposed hierarchical classification method. The following subsections present a discussion of the proposed hierarchical classification approach.

3.1. Detection of drivers engagement in secondary task

Applying the different classification methods considering only two classes results in a significant increase in classification accuracy. This is expected since the two classes are easily distinguishable when inspecting the PCA 2-D projection, which indicates that a good decision boundary can be constructed between the two classes. Table 4 shows the classification accuracy obtained by the different methods with 5-fold cross validation. The table shows that while some classification methods such as Gaussian NB can detect secondary task engagement (Class #1) with a considerably high accuracy, it has a poor detection accuracy for normal driving (Class #2), indicating high tendency for false positive detections. On the other hand, the Decision Tree method outperforms all other methods in detecting secondary task engagement with an outstanding accuracy of 99.5% and normal driving with an

Table 3
Supervised Classification Results (Four Classes).

Method	KNN	SVM Linear	SVM Nonlinear	Decision Tree	Random Forest	MLP	AdaBoost	Gaussian NB	QDA
Accuracy (%)	64.3	57.2	62.8	77.3	72.8	59.2	58.2	52.1	54.2

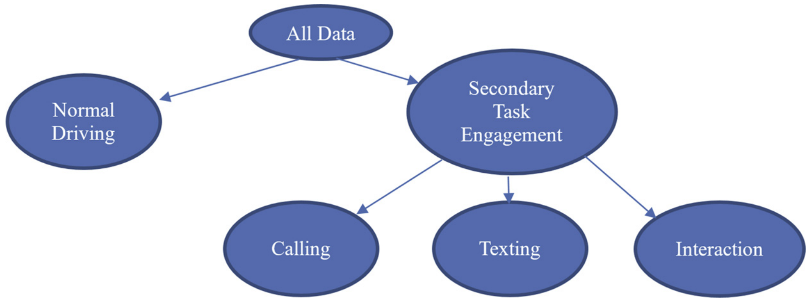

Fig. 3. The proposed hierarchical classification approach.

accuracy of 99.7%, resulting in an overall accuracy of 99.8%. This is followed by QDA method with an overall accuracy of 99.3% (Class# 1 accuracy = 99.4% and Class # 2 accuracy = 99.1%). Since the Decision Trees has considerably high detection accuracies compared to all other methods, it is further considered for error analysis.

3.1.1. Error analysis of decision tree

Decision trees (Dietterich, 2000) use feature values in the feature vectors to split the data into subsets. Each node in a decision tree corresponds to a split in the data. The splitting of data stops when all elements in every subset is assigned to a single class. One of the main advantages of decision trees is the ability to automatically reduce the dimensionality of the data during construction.

The classification errors associated with the two methods obtained for the training and testing datasets and shown in Fig. 4. As illustrated in Fig. 4, the training and testing errors converge quickly and almost linearly to zero as the max tree depth increases. The error analysis is done on a 5-fold cross validation to avoid overfitting. The results in Fig. 4 show almost the same convergence rate for both training and testing errors indicating that cross validation achieved its goal (Domingos, 2012).

Fig. 4. Training performance of the Decision Tree (Training and testing errors).

3.2. Identification of type of secondary task

Since identifying the types of the secondary tasks is not achieved with direct classification of the data, hierarchical classification is applied. In other words, instead of performing classification on four labels (classes), the data is first classified considering a binary classification model at one level, as discussed in the previous subsection. Then, at the second level, the data associated with secondary tasks is further classified using a multi-output classifier into three classes (hand-held cellphone calling “Class #1-1”, cellphone texting “Class #1-2”, and interaction with passengers “Class #1-3”). To determine the accuracy of secondary task identification, the accuracies obtained from the different levels of classification are multiplied. This hierarchical classification method has two advantages: (1) more detailed information on the data and multiple accuracy values are obtained instead of relying on a single accuracy value; and (2) using multiple classifiers helps obtain more reliable accuracy than a single classifier.

Table 5 shows the classification accuracy obtained for each secondary task type, as well as the overall accuracy of the second level of classification. The table shows that the Random Forest method outperforms all other methods in identifying the different types of secondary tasks with accuracy values of 77.2% for calling, 82.2% for texting, and 97.1% for interaction with passengers. The overall accuracy of differentiating between the three types is hence 82.2%. Although the Decision Tree method was able to detect drivers’ engagement in secondary tasks with an outstanding accuracy, as shown in Table 4, it did not maintain that performance when identifying the types of secondary tasks. One possible explanation for this outcome is the shape of the data illustrated in Fig. 2. As discussed earlier, this figure indicates that the points representing the three secondary tasks are not easily separable, thus a more complex classifier is needed.

3.3. Final model

The final classification model (or called secondary task identification) is illustrated in Fig. 5. In this model, as discussed earlier, the classification pipeline starts with the Decision Tree algorithm that solves a binary problem of whether drivers are engaged in secondary tasks or not using time-series vehicle kinematics data. Upon detection of drivers’ engagement in secondary tasks, the pipeline goes on to classify drivers further into three types depending on the secondary task they are engaged in (calling, texting, or conversation). The overall classification accuracy of this model is 82.03%, which is calculated as the multiplication of the accuracy obtained from the Decision Tree method at the first level by that of the Random Forest method at the second level (99.8 × 82.2). This accuracy is considerably acceptable, especially that the high accuracy of detecting drivers’ engagement in secondary tasks (99.8%) can help overcome risky driving behavior

Table 4
Supervised Classification Results (Two Classes).

Method	KNN	SVM Linear	SVM Nonlinear	Decision Tree	Random Forest	MLP	AdaBoost	Gaussian NB	QDA
Accuracy (%)	86.6	67.1	87.8	99.8	96.0	94.7	91.3	66.3	99.3
Class #1 Accuracy (%)	88.0	73.8	88.3	99.7	95.1	99.1	95.1	93.1	99.4
Class #2 Accuracy (%)	86.2	65.6	87.7	99.5	95.2	92.2	90.0	53.2	99.1

Table 5
Hierarchical Classification of Secondary Tasks.

Method	KNN	SVM Linear	SVM Nonlinear	Decision Tree	Random Forest	MLP	AdaBoost	Gaussian NB	QDA
Overall Accuracy (%)	66.8	60.1	64.4	79.7	82.2	57.6	60.0	55.5	55.9
Class #1-1 Accuracy (%)	37.2	7.2	20.1	73.0	77.2	1.1	3.1	2.1	3.1
Class #1-2 Accuracy (%)	33.2	6.0	16.3	76.8	82.2	1.0	3.0	6.2	6.0
Class #1-3 Accuracy (%)	91.5	91.1	94.9	85.1	97.1	97.7	99.5	94.2	95.1

resulting from such distractions. The complete results of the final model are provided in Table 6.

The final model, in Fig. 5, is investigated further to identify the features that impacted classification the most at each level of the hierarchy (secondary task detection and secondary task identification). Fig. 6 indicates that standard deviations of the speed parameter is the most critical feature in detecting secondary task engagement with an importance level of 47%. This is followed by the longitudinal acceleration with an importance level of 12%, which indicates that only the standard deviation of the speed feature is considered critical in detecting drivers' engagement in secondary tasks. Whereas, for identification of the secondary task types, almost all features have comparable importance levels ranging from 9% to 14%, except for the Yaw Rate which has an importance level of 7.5%. These results are expected to a far extent as the data were easily separable when comparing normal driving behavior to driving behavior associated with secondary task engagement (Fig. 2(b)). On the other hand, it was not as easy to separate the driving behavior associated with each type of secondary task (Fig. 2(a)). This is why almost all driving behavior performance measures along with their standard deviations were very important to extract distinct patterns associated with the different types of secondary tasks.

4. Conclusions

This study presents a methodology for identification of secondary tasks as one of the main causes of distracted driving behavior. To the authors' knowledge, past research has only investigated the possibility of detecting drivers' engagement in secondary tasks from observed driving behavior (Ye et al., 2017b,c). Yet, there has not been any attempt to identify the types of secondary tasks from driving behavior parameters. The proposed methodology in this study uses driving behavior parameters (speed, longitudinal acceleration, lateral acceleration, pedal position, and yaw rate) as well as their standard deviations to identify the different types of secondary tasks drivers are engaged in while driving. Three secondary tasks (identified in the literature as critical in changing driving behavior) are considered: hand-held cellphone calling, cellphone texting, and interaction with an adjacent passenger. Nine ensemble tree algorithms are used and comparatively evaluated for identification of secondary tasks. The algorithms include K Nearest Neighbor (KNN), Random Forest, Support Vector Machine (SVM), Decision Trees, Gaussian Neighborhood, Adaptive Boost (AdaBoost), Multilayer Perceptron (MLP), and Quadratic Discrimination Analysis (QDA).

The initial analysis showed that it is not easy to identify the types of secondary tasks by applying the different classification methods directly to the data. After further investigation of the data using unsupervised clustering, a hierarchical classification methodology was proposed. More specifically, a bi-level classification methodology was applied wherein the first level drivers' engagement in secondary tasks was detected, while in the second level, the secondary task engagement was further classified to identify the different types of secondary tasks. The main advantages of this hierarchical classification approach are that more detailed information on the data and accuracies at each level are obtained instead of relying on a single accuracy value, and also using multiple classifiers (one at each level) can help obtain more reliable accuracy than a single classifier.

The results showed that the Decision Tree classifier outperformed all other classifiers to detect drivers' engagement in secondary tasks with an outstanding accuracy of 99.8%. Whereas, the detection accuracy of all other classifiers ranged from 66% to 96%. When identifying the types of secondary tasks, the Random Forest classifier was found to be superior to all other classifiers with an overall accuracy of 82.2%, while all other classifiers had an overall accuracy between 55% and 79%. The accuracy of identifying each type of secondary tasks was 77.7% for hand-held cellphone calling, 82.2% for cellphone texting, and 97.1% for interaction with an adjacent passenger. These outstanding accuracy values might indicate that overfitting exists. However, the fact that these accuracy values are resulting from the testing datasets to which the models were not exposed during the training stage prove otherwise; this is in addition to the use of cross validation which helps avoid overfitting. These high accuracies indicate that drivers could have distinct driving behavior patterns, depending on the type of secondary task they perform. Hence, it is possible to use these distinct patterns as a means to identify the types of secondary tasks drivers engage in while driving.

In light of the research questions mentioned earlier and given the data used in the study, the research findings provide enough evidence that different secondary tasks may affect driving behavior differently which allowed for accurate identification of the different types of secondary tasks using the applied machine learning techniques. These promising findings can help identify counter measures to avoid safety

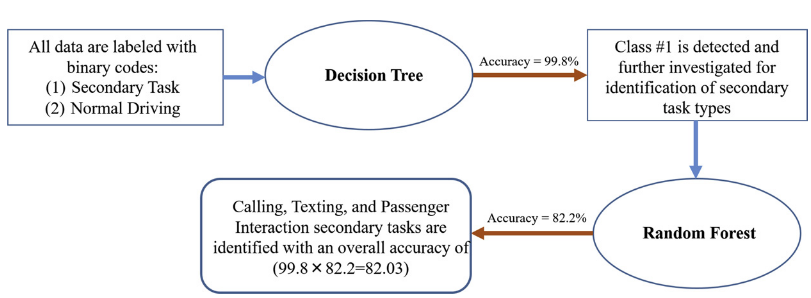

Fig. 5. The pipeline of the final classification model.

Table 6
Hyperparameters and Confusion Matrix of the Final Model.

	Hyperparameters	Confusion Matrix				
First Level (*Decision Tree*)	*Maximum Tree Depth* = 40			Predicted Classes		
				Secondary Task Engagement	*Normal Driving*	
		Actual Classes	*Secondary Task Engagement*	99.5%	0.5%	
			Normal Driving	0.3%	99.7%	
Second Level *(Random Forest)*	*Maximum Tree Depth* = 80 *No. of Trees* = 15 *Maximum No. of Features Considered when Looking at the best Split* = 3			Predicted Classes		
				Calling	*Texting*	*Interaction*
		Actual Classes	*Calling*	77.2%	2.0%	20.8%
			Texting	4.0%	82.2%	13.8%
			Interaction	1.9%	1.0%	97.1%

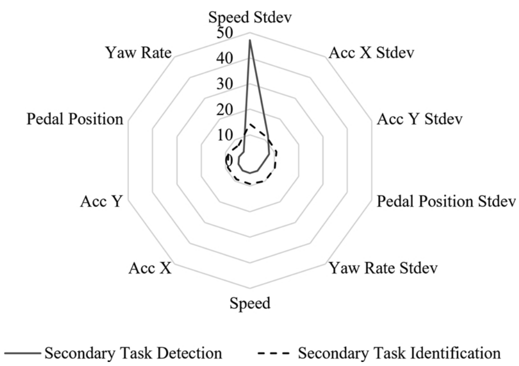

Fig. 6. Features importance level.

critical events associated with distracted driving behavior. With the technological advances of the emerging connected vehicles, the required data for the proposed model can be available. Hence, the proposed model can be applied for identification of whether drivers were engaged in unlawful secondary tasks in accident investigations for insurance claims purposes. Additionally, the proposed secondary task identification model can be incorporated into in-vehicle systems to detect risky driving behavior and alert drivers to pay attention back to the main task of driving.

It is worth pointing out that this study did not account for the effect of roadway type and geometric features and vehicle characteristics on the driving behavior variables. However, the driving behavior variables are analyzed as a pattern recognition problem in this study. In other words, identification of secondary tasks is performed through studying the pattern of changes in the driving behavior variables, rather than targeting specific values of each variable as indicators of the type of secondary task drivers are engaged in. Nonetheless, future research will study the impact of roadway type and geometric features and vehicle characteristics on driving behavior variables, hence on the predictability power of the developed models.

Acknowledgment

We would like to acknowledge Misagh Naderi for his assistance with the data extraction task.

References

Ahmed, M.M., Ghasemzadeh, A., 2017. Exploring the impacts of adverse weather conditions on speed and headway behaviors using the SHRP2 naturalistic driving study data. In: the 96th Annual Meeting of the Transportation Research Board. Washington D.C.

Anon, 2013. The 2nd Strategic Highway Research Program Naturalistic Driving Study Dataset. https://doi.org/10.15787/VTT1D59V.

Anon, 2014. Traffic Safety Facts 2012 Data: Older Population. National Highway Traffic Safety Administration.

Ashley, G., Osman, O.A., Ishak, S., Codjoe, J., 2017. Effect of secondary tasks and driver behavior on crash and near-crash risk: naturalistic driving study. In: the 96th Annual Meeting of the Transportation Research Board. Washington D.C.

Bagdadi, O., 2013. Assessing safety critical braking events in naturalistic driving studies. Transp. Res. Part F Traffic Psychol. Behav. 16, 117–126.

Bakhit, P.R., Osman, O.A., Ishak, S., 2017. Detecting imminent lane change maneuvers in connected vehicle environments. Transp. Res. Rec.: J. Transp. Res. Board 168–175 No. 2645.

Chang, L.Y., Chien, J.-T., 2013. Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. Saf. Sci. 51 (1), 17–22.

Consiglio, W., Driscoll, P., Witte, M., Berg, W.P., 2003. Effect of cellular telephone conversations and other potential interference on reaction time in a braking response. Accid. Anal. Prev. 35 (No. 4), 495–500.

Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. Mach. Learn. 40 (No. 2), 139–157.

Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J., Perez, M., Hankey, J., Ramsey, D., Gupta, S., 2006. The 100-car Naturalistic Driving Study, Phase II-results of the 100-car Field Experiment. Virginia Tech Transportation Institute.

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. Proc. Natl. Acad. Sci. 113 (10), 2636–2641.

Domingos, P., 2012. A few useful things to know about machine learning. Commun. ACM 55 (10), 78–87.

Ghasemzadeh, A., Ahmed, M.M., 2017. Driver's lane keeping ability in heavy Rain: preliminary investigation using the SHRP2 naturalistic driving study data. In: the 96th Annual Meeting of the Transportation Research Board. Washington D.C.

Guo, F., Fang, Y., 2013. Individual driver risk assessment using naturalistic driving data. Accid. Anal. Prev. 61, 3–9.

Harb, R., Yan, X., Radwan, E., Su, X., 2009. Exploring precrash maneuvers using classification trees and random forests. Accid. Anal. Prev. 41 (1), 98–107.

Hofleitner, A., Herring, R., Bayen, A., 2012. Arterial travel time forecast with streaming data: a hybrid approach of flow modeling and machine learning. Transp. Res. Part B Methodol. 46 (No. 9), 1097–1122.

Horrey, W.J., Wickens, C.D., 2004. Cell phones and driving performance: a meta-analysis. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, No. 48. SAGE Publications Sage CA, Los Angeles, CA, pp. 2304–2308.

Hosking, S.G., Young, K., Regan, M.A., 2005. The Effects of Text Messaging on Young Novice Driver Performance.

Jenkins, S., Codjoe, J., Alecsandru, C., Ishak, S., 2017. Exploration of the SHRP 2 NDS: development of a distracted driving prediction model. Advances in Human Aspects of Transportation. Springer, pp. 231–242.

Jonasson, J.K., Rootzén, H., 2014. Internal validation of near-crashes in naturalistic driving studies: a continuous and multivariate approach. Accid. Anal. Prev. 62, 102–109.

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.

Knipling, R.R., 2017. Crash heterogeneity: implications for naturalistic driving studies and for understanding crash risks. In: In the 96th Annual Meeting of the Transportation Research Board. Washington D.C.

Li, Y.E., Gibbons, R.B., Medina, A., Bhagavathula, R., Milton, J., Van Schalkwyk, I., 2017a. Correlation between roadway lighting and driver behavior at freeway interchanges based on naturalistic driving data. In: the 96th Annual Meeting of the Transportation Research Board. Washington D.C.

Li, Y., Medina, A., Gibbons, R.B., 2017b. Using naturalistic driving data for assessing safety impact of roadway improvements. In: the 96th Annual Meeting of the Transportation Research Board. Washington D.C.

McEvoy, S.P., Stevenson, M.R., McCartt, A.T., Woodward, M., Haworth, C., Palamara, P., Cercarelli, R., 2005. Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. BMJ 331 (No. 7514) p. 428.

Moazenzadeh, R., Mohammadi, B., Shamshirband, S., Chau, K.-w., 2018. Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran. Eng. Appl. Comput. Fluid Mech. 12 (No. 1), 584–597.

Mousa, S.R., Bakhit, P.R., Osman, O.A., Ishak, S., 2018. A comparative analysis of tree-based ensemble methods for detecting imminent lane change maneuvers in

connected vehicle environments. Transp. Res. Rec p. 0361198118780204.
Nagalla, R., Pothuganti, P., Pawar, D.S., 2017. Analyzing Gap Acceptance Behavior at Unsignalized Intersections Using Support Vector Machines, Decision Tree and Random Forests. Procedia Comput. Sci. 109 (Supplement C), 474–481.
Osman, O.A., Hajij, M., Karbalaieali, S., Ishak, S., 2018. Crash and near-crash prediction from vehicle kinematics data: a shrp2 naturalistic driving study. In: the 97th Annual Meeting of the Transportation Research Board. Washington D.C.
Sangster, J., Rakha, H., Du, J., 2013. Application of naturalistic driving data to modeling of driver car-following behavior. Transp. Res. Rec.: J. Transp. Res. Board 20–33 No. 2390.
Smith, B.M., Dyer, C.R., Chitturi, M.V., Lee, J.D., 2017. Automatic driver head State estimation in challenging naturalistic driving videos. In: the 96th Annual Meeting of the Transportation Research Board. Washington D.C.
Stutts, J., Feaganes, J., Rodgman, E., Hamlett, C., Reinfurt, D., Gish, K., Mercadante, M., Staplin, L., 2003. The causes and consequences of distraction in everyday driving. Annual Proceedings/Association for the Advancement of Automotive Medicine, No. 47. Association for the Advancement of Automotive Medicine p. 235.
Stutts, J., Feaganes, J., Reinfurt, D., Rodgman, E., Hamlett, C., Gish, K., Staplin, L., 2005. Driver's exposure to distractions in their natural driving environment. Accid. Anal. Prev. 37 (No. 6), 1093–1101.
Sun, Y., Xu, H., Wu, J., Hajj, E.Y., Geng, X., 2017. Data processing framework for driving cycle development with naturalistic driving study data. In: the 96th Annual Meeting of the Transportation Research Board. Washington D.C.
Vanajakshi, L., Rilett, L.R., 2007. Support vector machine technique for the short term prediction of travel time.In intelligent vehicles symposium. 2007 IEEE. IEEE, pp. 600–605.
Victor, T., Dozza, M., Bärgman, J., Boda, C.-N., Engström, J., Flannagan, C., Lee, J.D., Markkula, G., 2015. Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk.
Weiss, S.M., Kulikowski, C.A., 1991. Computer Systems That Learn: Classification and Prediction Methods From Statistics, Neural Nets, Machine Learning, and Expert Systems. Morgan Kaufmann Publishers Inc.
Weng, J., Meng, Q., 2012. Ensemble tree approach to estimating work zone capacity. Transp. Res. Rec.: J. Transp. Res. Board 56–67 No. 2286.
Wong, J.-T., Huang, S.-H., 2013. Attention allocation patterns in naturalistic driving. Accid. Anal. Prev. 58, 140–147.
Ye, M., Osman, O.A., Ishak, S., 2017a. Accounting for driver distraction and socioeconomic characteristics in a crash risk index: naturalistic driving study. Transp. Res. Rec.: J. Transp. Res. Board (2659), 204–211.
Ye, M., Osman, O.A., Ishak, S., Hashemi, B., 2017b. An ANN model for detecting secondary tasks from driving behavior attributes: a naturalistic driving study. In: Presented at the 96th Annual Meeting of the Transportation Research Board. Washington D.C.
Ye, M., Osman, O.A., Ishak, S., Hashemi, B., 2017c. Detection of driver engagement in secondary tasks from observed naturalistic driving behavior. Accid. Anal. Prev. 106, 385–391.
Zhang, Y., Haghani, A., 2015. A gradient boosting method to improve travel time prediction. Transp. Res. Part C Emerg. Technol. 58, 308–324.