Classification of Officers’ Driving Situations Based on Eye-Tracking and Driver Performance Measures

Maryam Zahabi, Member, IEEE, Yinsong Wang, and Shahin Shahrampour, Senior Member, IEEE

Abstract—Motor vehicle crashes are a leading cause of police officers’ deaths in the line of duty. These crashes are mainly attributed to officers’ use of in-vehicle technologies while driving, distraction, fatigue, and high-speed driving conditions. The objective of this study is to classify officers’ driving situations using a combination of driver behavior and eye-tracking measures. The study compared three algorithms, including random forest (RF), support vector machine (SVM), and random Fourier features (RFF) to classify officers’ driving situations (i.e., normal vs. pursuit driving) and in-vehicle technology use. The results suggested that driver behavior measures, combined with RF or SVM methods, are most promising for classifying officers’ driving condition (accuracy of about 90%). However, it might be more efficient to apply RFF with driver behavior measures to classify officers’ use of in-vehicle technologies while driving due to the time cost reduction of RFF as compared to SVM and RF algorithms. The findings can be applied to improve future police vehicles, training protocols, and to provide adaptive technology solutions to reduce officers’ driving distraction and workload.

Index Terms—Classification, driving, police, random forest (RF), random Fourier features (RFF), support vector machine (SVM).

I. INTRODUCTION

MOTOR vehicle crashes are a leading cause of police officers’ deaths in the line of duty [1]. About 1600 officers have been killed due to motor vehicle crashes in the past decade [1]. The number of crashes is significantly higher than the crashes involving fire trucks and ambulances [2] and about 2.5 times higher than the national average for other occupations [3]. Police vehicle crashes have been mainly attributed to officers’ use of in-vehicle technologies while driving, distraction, fatigue, high speed driving conditions including emergency and pursuit situations, and individual characteristics such as age and experience [4]–[6].

Prior studies have been mainly focused on assessing officers’ driving performance, workload, and distraction due to the use of in-vehicle technologies while driving [7]–[9]. Police vehicles are equipped with several technologies such as the mobile computer terminal (MCT), radio, cell phone, radar, siren and control panel, and video cameras (see Fig. 1). The MCT has been found to be the most important and frequently used technology for police officers [9]. The MCT can increase officers’ productivity by providing real-time information and reduce the amount of paperwork [10]; however, it is not designed to use while the vehicle is in motion [11]. Previous investigations have found a number of usability and interface design issues with the MCT and, therefore, the use of this technology while driving has increased officer distraction and workload [7], [8], [12]. Some of the MCT interface usability issues identified from previous studies included poor filtering and layout of information presentation, delayed interaction, interface not optimized for touch screen, and issues with navigation and routing [10], [13], [14]. However, officers frequently use in-vehicle technologies, especially the MCT while driving as part of their job [15]. Observations of police daily activities revealed that about 77% of officers used MCTs while driving, 55% of officers used MCTs while performing at least one other task (e.g., using the radio or cell phone), and 11% used MCTs while performing at least two other tasks simultaneously [15].

Some recent studies investigated the effect of driving conditions on officers’ performance and workload. For example, Shupsky et al.’s study [7] comparing the operational (e.g., lane keeping, following the lead car) and tactical (e.g., passing maneuvers or overtaking) driving conditions indicated that officers exhibited better driving (in terms of steering entropy and speed deviation) but worse secondary task performance under the operational driving as compared to the tactical driving condition. In addition, in our recent study [16], it was found that officers exhibited higher level of cognitive load in the pursuit driving...
condition as compared to normal driving. Higher demands in the pursuit driving condition are mainly due to the need for officers to drive at high speed, change lanes, communicate with the dispatch or other officers via radio or MCT to get information about the case, and perform sudden maneuvers as compared to the normal condition.

Some prior research provided adaptive MCT interfaces based on officers’ work context or preferences. For example, Kurkinen et al. [17] developed an interface, which recognized the cognitive demands of officers while driving based on the vehicle’s location, speed, and pursuit situation (light/siren activation) and could adapt by using a combination of symbolic and text-based information or capturing critical information automatically. In another study, Streefkerk et al. [18] developed a design concept for an adaptive MCT that could adjust the information presentation format based on the driver’s cognitive capacity, tasks, context, or individual preferences. However, their work was limited to a design concept and requires statistical modeling techniques to predict officer behavior and tasks in real-time. With recent advances in wearable devices, sensors, and state-of-the-art machine learning (ML) techniques, police in-vehicle technologies such as the MCT can be designed to adapt in real-time and provide alerts when the officer is under high workload conditions or distracted. However, to achieve this goal, there is a need to use modern classification techniques that incur small computational cost, making them well-suited for time-sensitive applications.

Previous studies used ML algorithms to classify driver distraction or workload using a combination of driver behavior and physiological measures. For example, in a recent study, McDonald et al. [19] compared several ML algorithms such as random forest (RF) and support vector machines (SVM) and a combination of physiological and driver behavior inputs to classify driver distraction. It was found that the RF algorithm trained by using only driver behavior measures and excluding physiological measures was the best algorithm to classify driver distraction. Solovey et al. [20] used a combination of physiological measures (electrocardiogram (ECG), skin conductance) and driver behavior responses (speed, steering wheel position) and compared several classification algorithms such as logistic regression and decision trees to classify driver workload and evaluate the in-vehicle user interfaces. The findings indicated that using multimodal inputs including both physiological and driver behavior measures led to the highest algorithm accuracy (around 90%). In another study [21] and using the SVM algorithm, driver behavior, physiological, and eye-tracking measures to classify driver workload, it was found that the SVM could classify driver workload with an accuracy of around 89% using lane deviation, physiological, and gaze data.

Other studies used single modal features such as driver behavior or physiological measures for classification of driver workload. For example, Son and Park [22] found that driver workload can be classified with an accuracy of about 73% using only the lane position and steering wheel reversal rate. In another study, Tjolleng et al. [23] used only ECG inputs and classified driver workload with 82% accuracy using the artificial neural network algorithm.

A. Motivation

Although prior studies have identified the negative impacts of police in-vehicle technology use and driving conditions on officers’ workload, performance, and distraction [7]–[9], there has been limited effort on detection and classification of officers’ driving situations (e.g., whether the officer is involved in a pursuit situation or is engaged in secondary tasks while driving). Furthermore, previous investigations on driver distraction and workload classification were limited to civilian drivers under normal driving situations. Technologies in police vehicles are more complex than the equipment in civilian vehicles and they require greater cognitive demands. Also, the officers are under high stress and workload in pursuit driving situations [24] and the severity of their crashes is higher than the severity of crashes involving civilian drivers [25].

Real-time detection and classification of officers’ driving situations is critical and has a potential to improve officers’ driving safety by providing alerts or adaptive automation features in the vehicle. In addition, due to the time sensitivity of this application, there is a need to use efficient ML algorithms, which incur small computational cost. Therefore, the objective of this study was to provide an efficient ML algorithm to classify officers’ driving situations (i.e., whether the officer is involved in a pursuit or normal situation or is using in-vehicle technologies while driving) using a combination of eye-tracking and driver behavior measures.

II. METHOD

A. Data Collection

The dataset used in this analysis was collected using the STISIM fixed-based driving simulator setup (System Technology, Inc., Hawthorne, CA) located at Texas A&M University (see Fig. 2). The eye-tracking measures were collected using the Pupil Labs eye-tracking glasses including a world camera with 100 degrees field of view and two eye cameras (to capture pupillometry data). Eighteen police officers (age: $M = 36.29$ yrs., $SD = 6.52$ yrs.) from different departments in Texas participated in the experiment. All participants were driving police vehicles on a regular basis (average of 14 h. per shift), had high experience using in-vehicle technologies such as MCT while driving (average technology experience of 82.34%). In addition, all participants except one had prior experience driving in pursuit situations.

1) Study Process: The study simulated four secondary task situations including no secondary task, single secondary task (license plate number check task with the MCT or radio communication task), and multiple secondary tasks (MCT and radio communication tasks simultaneously). Each experimental trial simulated a pursuit or normal driving condition. The experiment followed a within-subject design (4 secondary task × 2 driving condition) with replication and, therefore, the study included 16 experimental trials. In each experimental trial, two blocks of data collection existed and therefore, the total number of data points or measurement epochs were 576 (i.e., $18 \times 4 \times 2 \times 2 = 576$). In trials including secondary tasks, the officers were
asked to perform the secondary tasks in both data blocks. In trials without any secondary task, the officers were only performing the primary task (i.e., driving) in both data blocks.

All participants experienced the driving trials in a random order. Prior to the experiment, participants filled out the consent form and the background questionnaire. They were provided with three training scenarios to become familiar with the driving simulator controls. After the initial familiarization with the driving simulator, they were provided with the instructions regarding the secondary tasks (i.e., MCT and radio communication tasks). Once the participants were comfortable with the secondary tasks, the eye-tracking glasses were calibrated to participants’ eyes and they started the actual experiment. Each driving trial took approximately 3 min. to complete and the participants were provided with a 2-min break between trials. After the completion of experimental trials, participants were debriefed and compensated $60 for their time. The experiment took approximately 2 h to complete. The Texas A&M institutional review board reviewed and accepted the study procedure.

2) Driving Simulation Scenario: The driving scenarios simulated an urban environment with moderate traffic density. They presented a six-lane roadway condition with three lanes on either side of a double-yellow line with opposite directions of travel. Each scenario included four intersections. To increase the realism of the pursuit driving condition, the lead vehicle changed its lane several times and the officers had to follow the path and made sudden lane changes. Furthermore, having a moderate traffic density level and several intersections made the process of car following more challenging for the officer to avoid crashes, which is similar to actual police pursuit situations.

3) Secondary Tasks: In scenarios including the secondary tasks, in two randomly selected locations along the drive, the officers were asked to perform the license plate number check task and/or listen to a radio communication. For the plate number check task, an automated voice from the simulator provided the question (e.g., “what is the vehicle insurance status?”). The questions were designed based on our prior studies and interviews with police officers [9]. Participants then searched for the information on the MCT (by pressing the arrow keys to go to different information pages and reading the information on each page). The task completed once the officer verbally provided the answer to the experimenter. Subsequently, an automated voice from the simulator provided a question regarding that radio communication (e.g., “What is the suspect wearing?”). The task completed once the officer verbally provided the answer to the experimenter.

4) Dataset: Driver behavior and eye-tracking measures were collected continuously throughout each driving trial. Driver behavior measures included speed and lane deviation responses. In the normal driving condition, the speed deviation response was defined as the absolute deviation of the vehicle’s speed from the posted speed limit (40 mph). The lane deviation response was defined as the absolute lateral deviation between the center of the vehicle and the center of a lane. In the pursuit condition, speed deviation was calculated as the absolute deviation of the vehicle’s speed from 65 mph. Since lane changes can impact speed deviation, the data collection blocks captured speed deviations in areas where the officers did not change lanes. Furthermore, since the officer was asked to follow the lead vehicle, the lane deviation was defined as the absolute lateral deviation between the center of the vehicle and the lead vehicle. Although the lane deviation definitions were different between the normal and pursuit driving conditions, in both situations, officers had to follow some “instructions” (i.e., normal driving condition: “stay in the middle lane at all times,” pursuit driving condition: “follow the path of the lead vehicle”) and deviations from those “instructions” was defined as lane deviation. Driver behavior measures of speed and lane deviation have been used in prior research assessing driver workload, visual and/or cognitive distraction, and as input measures for classification algorithms [26], [27].
The eye-tracking measures included the percentage change in pupil size (PCPS) and blink rate. The PCPS measure was calculated as the percentage of difference between the average pupil size in each data collection block (e.g., when the officer was interacting with the MCT and/or radio) and the baseline pupil size (collected prior to the experiment when the officers were asked to relax while seated in the cab) divided by the baseline pupil size. These measures have been used in prior studies including driving simulation studies as measures of cognitive workload [28], [29]. The driving and eye-tracking measures were collected with a sampling frequency of 60 Hz.

B. Data Preprocessing

The complete dataset included 576 data points. Thirteen data points in the speed deviation response were identified as outliers and were removed, since the absolute speed deviation in these cases were more than 10 mph, which indicated that the officers were not following the instructions. In addition, 43 data points in the lane deviation response were removed in cases that the officers changed their lane and did not follow the instructions (e.g., stay in the middle lane at all times). Furthermore, 51 blink rate data points, 113 pupil size data points for the right eye and 155 data points for the left eye had to be removed due to the issues with the eye-tracking glasses such as calibration problems or glare from the screen. However, the missing data were approximated and substituted using multiple imputation with chained equations (MICE) approach [30]. The dataset was preprocessed using Python through three steps, including one-hot encoding, MICE, and training testing split.

1) One-Hot Encoding: Although both modes of data (driver behavior measures and eye-tracking measures) fed into ML algorithms are numerical, it is still necessary to utilize the categorical information such as participant number, trial number, and replications. For example, the blink rate might increase as a function of time on task [31]. This assumption was backed up by the observation that among all the participants with their blink rate data successfully recorded, 10 participants showed their highest blink rate during the second replication. Moreover, the majority of the missing data came from particular participants (e.g., due to the glare or calibration issues), therefore, it is important to take participant number into account during the missing data handling phase.

To utilize the categorical information in an unbiased way in the regression-based handling of missing data, one-hot encoding was adopted in this study. One-hot encoding, available in scikit-learn [32], turns a multiclass categorical attribute into several binary (0-1) attributes, where each new attribute indicates whether a data point is in this class or not. By doing this, all new attributes can be used in a regression model (to categorize whether a data point is in this class or not). By doing this, all new attributes can be used in a regression model (to categorize whether a data point is in this class or not). By doing this, all new attributes can be used in a regression model (to categorize whether a data point is in this class or not).

2) MICE Equations: A large portion of the data points were not available directly for both training and testing. However, a data point with missing entries might still contain other attribute values that might be useful in model building. Therefore, we did not simply disregard these points. To utilize as much information provided by the dataset as possible, we adopted MICE to fill the missing entries.

MICE iteratively fit regression models to fill out the missing entries. More specifically, for the five attributes (i.e., lane deviation, speed deviation, PCPS for the left eye, PCPS for the right eye, and blink rate) with missing entries in this dataset, MICE first fills the missing entries by the mean value of the available entries; then, it iteratively builds a Bayesian linear regression model over these attributes to complete the process. After excluding the driving condition and secondary tasks attributes, for each attribute with missing entries, data points without missing entries in this attribute will be used as the training set to build a linear regression with this attribute as the output and the other attributes (including one-hot encoded categorical attributes) as the input. Then, the missing entries will be updated as the test output of the Bayesian linear regression model, which could account for uncertainty. This iteration will be repeated multiple times. In this study, the number of iterations was set to be 50, which could update all 5 attributes 10 times each. The whole process was repeated 40 times to create 40 different imputed datasets. In the simulation, the classification accuracy of an algorithm was evaluated as the average accuracy of these 40 datasets. A flowchart for MICE is presented in Fig. 3.

\[ D_s (s = 1, ..., S) \] are the imputed datasets. \( m_i \) and \( n_i \) are the index for nonmissing entries and missing entries at the ith attribute (i = 1, ..., d). m contains \( m_1 \) to \( m_d \), n contains \( n_1 \) to \( n_d \) and \( t = 1, ..., T \) is the index for iterates.

3) Dataset Training and Testing: Due to the allocation of missing data, it is necessary to split the training and testing
data randomly meanwhile maintaining a ground truth input. All data points that originally had missing entries were excluded in the testing data selection to avoid multiple random inputs due to MICE procedure. After removing the missing data, 60 data points were sampled randomly from the remaining 341 data points as the testing dataset. The other 516 data points were served as the training set. Each training and testing data split was used for all 40 imputation dataset in one simulation.

C. Algorithm Training and Evaluation

The classification models for both scenarios were constructed with three ML algorithms. In addition to the two baseline algorithms RF and SVM compared by McDonald et al. [19], an approximation algorithm, Random Fourier Features (RFF), was also included in the baselines. RFF reduces the computation time of kernel SVM using kernel approximation at the cost of some classification accuracy. The RFF was introduced by Rahimi and Recht [33] to alleviate the computation complexity issue raised by kernel-based ML algorithms, including Kernel Ridge Regression, Kernel Logistic Regression, and one of the baseline algorithm SVM. Despite their success in achieving good classification accuracy, kernel-based algorithms suffer from a prohibitive training cost, which scales at least quadratically (and more often cubically) with respect to the number of data points used in training. The time complexity of evaluating these algorithms at a test point also scales linearly with respect to the number of training data. In contrast, RFF can approximate any kernel-based algorithms meanwhile maintaining a linear training complexity and a test complexity that is independent of the number of training data. In the context of this study, ML algorithms are expected to work in real time under the condition where officers might be in the pursuit of another vehicle or under high demand situations performing secondary tasks. Therefore, the time cost of the ML algorithm is critical, and the purpose of including RFF in this study can be naturally justified. As we will see, RFF shows great potential in this context, which will be discussed in detail in Section III.

The RF model was built with five random trees due to the small number of attributes in the dataset. To be specific, each imputed dataset generated one RF model with its own five random trees. The SVM model was built with radial kernel, where tuning parameters (kernel width and cost parameter) were identified with an exhaustive grid search. The grid search was conducted for each imputed dataset. To approximate the SVM used in this study, RFF also uses the same kernel width and cost parameter found by the grid search of SVM. The number of random features was set to be 20, which provides a decent performance while maintaining its computational advantage.

III. RESULTS

A. Driving Condition Classification

The data preprocessing as well as algorithm training and evaluation were performed 50 times to address the uncertainty within the training testing splitting. Three types of input were used, including eye-tracking measures, driver behavior measures, and combined measures, forming three different analyses.

The first row of Fig. 4 shows the classification accuracy and the standard error of the three benchmark algorithms for these three analyses. In the analysis that utilizes eye-tracking measures to predict the driving condition, SVM with radial kernel outperforms both RF and RFF. However, RF is on par with SVM in the analysis involving driver behavior measures. It was also observed that RFF is not particularly useful in this analysis, because the drop in the classification accuracy is quite significant.

Fig. 4 also indicates that predicting the driving condition using driver behavior measures is better than using only eye-tracking measures. However, there is no sign of clear advantage for the combined measures input.

It is important to note that algorithms perform differently because each of them uses a specific approach for approximating the relationship between measures and the classification target. For example, RF is based on approximating the target function using indicator functions, and SVM is based on using kernel functions. In this work, the obtained data is real-world, so the “true underlying target function” (ground truth) is not known. Therefore, it is not possible to know a priori which method dominates the others. In real-world applications, it is recommended to try different methods to see their capability in capturing the ground truth. Prior studies such as [19] also observed high variations in the accuracy of different ML algorithms such as RF and SVM when classifying driver distraction using physiological and/or driver behavior measures.

B. In-Vehicle Technology Use Classification

The simulation for in-vehicle technology use classification was also repeated 50 times for each of the three analyses involving different modes of input (i.e., eye-tracking measures, driver behavior measures, and combined measures).

The second row of Fig. 4 illustrates the performance and the standard error of the three benchmark algorithms in in-vehicle technology use classification analyses. It was found that SVM outperformed RF consistently over all three analyses. However, unlike driving condition prediction, RFF shows an impressive performance in in-vehicle technology use classification; it achieves a classification accuracy that is ∼7% higher than RF when using driver behavior measures, while losing around 3% of accuracy in the rest of the tasks. In general, SVM with radial kernel appeared to be the best in terms of accuracy among the benchmark algorithms.

For the SVM and RF algorithms, the combined measures showed a slight advantage over the driving behavior or eye tracking measures (accuracy of around 77%–78%). However, RFF algorithm using only driver behavior measures yields statistically equivalent performance comparing to using combined measures (accuracy of 75%).

C. Time Cost Comparison

The training and testing time of the simulation were recorded and reported in Tables I and II. The training time of all three benchmark algorithms is evaluated for 40 imputed datasets and 50 replications, which amounts to a total of 2000 training cycles.
The testing time of all three benchmark algorithms is evaluated for 10,000 testing cycles on one imputed dataset.

Tables I and II illustrate that for a dataset with five attributes and 516 training data points, RF with five random trees takes the most time to train and test, SVM takes the least time to train and test, and SVM is in the middle. RF shows a 40% training time and 80% test time reduction compared to SVM and a 55% train time and 90% test time reduction compared to RF. This level of time advantage is achieved by RFF at the cost of less than 4% drop in the classification accuracy for in-vehicle technology use classification, suggesting that it is worthwhile to apply RFF in practice for this task. However, the performance difference in driving condition classification suggests that it might not be worthwhile to trade the performance drop for the time cost reduction.

In theory, the computation advantage of RFF is more visible as the number of training points increases. For example, tripling the amount of training points will lead to $27\times$ training time and $3\times$ testing time for SVM (order-wise). However, it will only triple the training time and maintain the same testing time for RFF. On the other hand, the training time of RF scales worse than random features as well, therefore, the run time difference between these two algorithms in percentage will become more significant as the number of training points increases.

### IV. DISCUSSION

Classification of officer driving situations is critical to provide adaptive technology solutions. Previous investigations on adaptive police in-vehicle technologies were limited to design concepts and prototypes, which have not been implemented in police vehicles [17], [18]. Furthermore, existing driver distraction and workload classification algorithms [19], [20] are developed based on civilian drivers’ performance and physiological measures in normal driving conditions, which might not be easily generalized to police driving conditions due to the differences among the police and private sector domains such as temporal demands placed on officers, travel speeds of vehicles, and the level of driver training [24]. The findings of this work can inform the use of efficient ML algorithms to classify officers’ driving conditions based on driver behavior or a combination of driver behavior and eye-tracking measures.
and to adapt in-vehicle technologies based on officers’ driving situations.

A. Driving Condition Classification

The results suggested that driver behavior measures were more promising in predicting the driving condition as compared to the eye-tracking measures. In addition, there was no significant advantage for the combined measures input (i.e., driver behavior and eye-tracking measures). These findings were expected based on the differences between the pursuit and normal driving conditions. The pursuit condition requires the officer to drive at high speed, change lanes, and perform sudden maneuvers. Furthermore, the officer is allowed to override the traffic regulations (e.g., speed limit) in pursuit conditions. However, in the normal driving condition, officers are required to follow all roadway regulations. Another explanation might be due to the differences between the normal and pursuit conditions in terms of driver level of control (i.e., operational, tactical, and strategic) [34]. The pursuit driving involves both operational (e.g., lane keeping, following the lead car) and tactical driving (e.g., passing maneuvers or overtaking), which is more demanding as compared to the normal condition that mainly requires operational driving behavior. Previous studies found officers to have worse driving performance (as indicated by speed variance) under operational driving as compared to the tactical driving condition [7]. Therefore, officer’s driving behavior measures including speed and lane deviation are beneficial in determining the driving condition. The findings are also in line with previous studies on civilian drivers that found driver behavior measures were the most important input measures for classifying driver distraction or workload [19], [21].

Driver behavior measures are also easier to acquire in real-time during police operations. Previous distraction and workload classification algorithms based on eye-tracking measures [35], [36] require real-time collection of those measures, which is comparatively more difficult and require additional in-vehicle data collection devices that are not commonly installed in regular vehicles [37]. Police vehicles are already overloaded with a number of in-vehicle equipment and adding an eye-tracking device might not be feasible. In addition, collecting eye-tracking measures such as blink rate and PCPS in realistic settings might be challenging due to the changes in the lighting condition and during night shifts [38].

The findings also suggested that for classifying driving condition, both RF and SVM algorithms provide high accuracy using driver behavior measures as inputs. These algorithms have been found useful for classification of workload and distraction in studies based on civilian drivers [19], [21], [27]. It is important to note that although the SVM was slightly more efficient (i.e., faster run time) as compared to the RF algorithm based on the simulations on the training and testing data points, future studies should validate this time cost advantage with larger datasets.

B. In-Vehicle Technology Use Classification

The findings on in-vehicle technology use classification suggested that both the RF and SVM achieved highest accuracy with combined driver behavior and eye-tracking measures. However, applying RFF with only driver behavior measures might be more efficient in this case due to the time cost reduction of RFF as compared to SVM and RF algorithms especially since the computation advantage of RFF will be even more visible with larger datasets. Furthermore, as mentioned earlier, driver behavior measures are easy to collect through readily available sensors and there is no need to implement additional equipment in the vehicle.

In-vehicle technology use (especially the MCT) while driving is one of the leading causes of police vehicle crashes [4]. Although several studies identified the effect of in-vehicle technologies on officers’ cognitive workload and distraction [7], [8], [39], there has been very limited effort on designing context aware solutions that can inform the officer of high-demand situations. Using the RFF algorithm with driver behavior measures, it might be possible to detect officers’ engagement in multitasking situations in real-time and provide timely alerts or automate some tasks to avoid cognitive overload and ultimately potential crashes.

The accuracy of RFF algorithm using driver behavior measures (around 75%) is also comparable with previous studies that used only vehicle-based measures to classify driver workload and distraction. For example, the SVM and RF algorithms with driver behavior input measures in McDonald’s study [19] classified driver distraction with accuracy of around 70%. In another study, Son and Park used the lane position and steering wheel reversal rate with radial basis probabilistic neural network and achieved driver cognitive workload classification accuracy of 73.3% [22]. Other investigations such as Jin et al. [40] achieved similar classification accuracies (approximately 74%) using driver behavior measures and the SVM algorithm. However, our results suggested that these algorithms are less efficient than the RFF algorithm. The findings of this study indicated that using the RFF algorithm is more beneficial in terms of “time-cost” despite being competitive in terms of “accuracy” as compared to the SVM and RF algorithms and therefore, can be more useful for classification of officers’ in-vehicle technology use which is a time sensitive task and should be performed in real-time.

C. Limitations and Future Work

The main limitation of this study was the small sample size due to the limited number of police officer participants and existence of missing data in the dataset due to eye-tracking calibration issues or participants not following the instructions. Future studies should evaluate the performance of algorithms proposed in this study with larger sample size. Second, the input measures used to train the algorithms in this study were limited. Although these measures are validated measures of driver behavior and workload and have been used extensively in the literature, other measures such as steering wheel angle, ECG, or skin conductance might be beneficial in classification of officers’ driving situations which need to be investigated further.

Furthermore, the experiment demonstrated the computation advantage of RFF by showing a 50% reduction in run time with
only 516 training points. It would be interesting to see how much gain in run time RFF will achieve in large dataset applications.

The other limitation was regarding the use of simulated driving scenarios and secondary tasks. Although the scenarios and secondary tasks were simulated based on actual police operations and in-vehicle technologies, and several prior studies validated the use of fixed-based driving simulators for assessing behavioral and physiological measures of driver behavior and distraction [41], [42], we should be cautious about generalizing the findings beyond the tasks and driving conditions explored in this study. Future research should evaluate the performance of proposed algorithms using naturalistic driving studies or ride-along and assess the generalization of the models with other secondary tasks (e.g., cell phone use) not included in the training of the models. In addition, the participants in this study perceived the pursuit driving condition (average rating of 2.84 out of 5) to be more cognitively demanding as compared to the normal driving condition (average rating of 2.05 out of 5). However, there is a need for future studies to assess how the simulated normal driving and pursuit driving conditions correspond to the types of workload that the officers might experience in the field.

Finally, RFF is only one kernel approximation technique that can be used to speed up kernel SVM. There exist a number of other kernel approximation techniques in the ML literature and all of these methods reduce the time cost of kernel SVM (perhaps) at the cost of some degradation in classification performance. Studying these techniques in this context is another interesting research direction.

D. Application

The findings of this study can be applied to improve future police vehicles and provide adaptive technologies to reduce officers’ driving distraction and workload. For example, the driving condition algorithm can be used to determine whether the officers are involved in pursuit situations and adapt the in-vehicle technology information presentation format accordingly (e.g., change the MCT display to low clutter mode, present the most important information, use of speech-based data entry format or head-up displays).

The secondary task algorithm can be beneficial in identifying whether the officer is using any secondary tasks while driving and can provide real-time alerts in cases that officers are involved in multitasking situations or are under high workload. The algorithms can also be beneficial for advancing police training protocols for high-demand situations involving pursuit driving and multitasking by allowing the officers to practice management skills and prioritize among task components. Most of the police agencies use San Jose field and evaluation training protocol, which is typically based on classroom training of MCT functionalities and evaluated based on subjective ratings, and require the officers to learn in the field how to balance between driving and managing secondary tasks. Implementing the proposed algorithms in police vehicles can be helpful in teaching officers how to allocate their attention to driving vs. secondary tasks in different driving conditions (e.g., by providing real-time feedback).

Compared to the findings of civilian driver studies, this study indicated that driving behavior responses are more promising in predicting officers’ driving situations as compared to eye-tracking measures. One possible explanation to this finding might be due to the complexity and necessity of in-vehicle technologies for police officers. While driving, the officers need to perform several tasks with the MCT (e.g., checking driver’s license and plate information, checking vehicle’s history) and communicate with the dispatch frequently to access real-time information. However, the in-vehicle technologies for civilian drivers are for performing routine activities (e.g., entertainment, navigation) and less demanding, and the driver is not required to perform these tasks while driving. Therefore, the use of these technologies while driving might degrade officer’s driving performance more compared to the use of in-vehicle technologies in civilian vehicles.

V. CONCLUSION

The objective of this study was to classify officers’ driving situations using a combination of driver behavior and eye-tracking measures. The findings suggested that driver behavior measures, combined with RF or SVM methods, are most promising (accuracy of about 90%) for classifying officers’ driving condition (normal vs. pursuit). However, it might be more practical to apply RFF with driver behavior measures to classify officers’ use of in-vehicle technologies while driving due to the time cost reduction of RFF as compared to SVM and RF algorithms. All three models should work with different types of data but we do recommend standardizing the data before implementing the algorithms.

ACKNOWLEDGMENT

The views and opinions expressed are those of authors and do not necessarily reflect the views of the OSHERC. The authors would like to thank Vanessa Nasr and Ashiq Mohammed Abdul Razak for their help with the data collection.

REFERENCES

9


