

A Cloud-Based Environment-Aware Driver Profiling Framework Using Ensemble Supervised Learning

Abdalla Abdelrahman
Electrical and Computer Engineering
Queen's University
Kingston, Ontario
Email: a.abdelrahman@queensu.ca

Hossam S. Hassanein
School of Computing
Queen's University
Kingston, Ontario
Email: hossam@cs.queensu.ca

Najah Abu-Ali
College of Information Technology
United Arab Emirates University
Al-Ain, UAE
E-mail: najah@uaeu.ac.ae

Abstract—Driver profiling is an emerging scheme that has a wide range of applications in the field of Intelligent Transportation Systems (ITS). Driver profiling is the real-time process of detecting driving behaviors and computing a driver's competence level based on detected behaviors. In this paper, a novel driver profiling framework is presented. A risk prediction model is hosted in the cloud to determine the risk associated with detected behaviors in specific driving environments. Risk values along with a driver's compliance to warnings are both utilized to compute a driver's risk profile. Using SHRP2 large-scale Naturalistic Driving (ND) dataset, the development of the risk prediction model is presented herein with the underlying sub-processes of data preprocessing, error analysis, and model selection. Validation results show that a developed randomized trees supervised learning model is proven to have a good trade-off between bias and variance with evidently high performance results.

Index Terms—Driver profiling, Intelligent Transportation Systems (ITS), driving behavior classification, supervised learning, random forests, telematics

I. INTRODUCTION

The recent advancements in vehicular sensing, cellular communications, as well as cloud computing have enabled the deployment of various Intelligent Transportation Systems' (ITS) applications. Given the high vehicle crash rates [1], these applications have the potential to considerably lower these rates.

An emerging safety-based ITS application is driver behavior profiling [2]. Driver profiling is the process of acquiring real-time vehicular data using CAN-bus through OBD II units or mobile-sensed data using inertial smartphone sensors to detect behaviors and warn drivers if risky behaviors are detected. Driver profiling has been widely deployed in different safety based applications. Pay-How-You-Drive (PHYD) is an example of car telematics insurance scheme in which an insurance premium is rated according to a driver's per-trip driving score [3]. Other emerging driver profiling applications include fleet telematics profiling systems [4], safety-based route planning, and driver self-coaching systems [5].

Most literature work in the context of driver profiling has been focusing on the detection of certain behaviors, which are considered risky. Detected behaviors are then inputted to scoring functions that assign different weights to detected behaviors based on the expected risk of each [6]. Not only

are such scoring functions subjective due to the absence of a valid risk measure (i.e., a risk measure quantified in terms of the actual risky events such as crash and near crash events), but also they ignored the environmental effect on risk given the detected behaviors. Naturalistic driving studies (NDSs) have provided large scale data about behavioral causes of risky events (i.e., crashes and near crashes), as well as the environmental context of such behaviors (e.g., weather and road conditions, traffic density level, etc.). In addition, NDSs provide the behavioral and environmental information during normal driving episodes, which enables the development of environmental-aware risk prediction models [7].

Recently, the Virginia Tech Transportation Institute (VTTI) conducted the largest NDS to date named SHRP2 NDS [7]. This dataset contains the behavioral and environmental contextual information of nearly 9,000 crash and near crash events and more than 20,000 events captured during normal driving episodes. The research question this paper is addressing is:

Are driving behavioral habits together with their environmental context good predictors for measuring risk probability?

To answer this question, the behavioral and environmental details of driving events presented in SHRP2 NDS are utilized to build a risk prediction model that can be incorporated in a complete cloud-based driver profiling framework as well be detailed in section III. The research contributions of this paper are summarized as follows:

- 1) An envisioned novel Cloud-based Environment-aware Driver Profiling (CEDP) system is presented and thoroughly discussed. The system provides a view on a “next generation” driver profiling system in which drivers are mainly profiled based on the statistical correlation between their behaviors and the actual risk probability given the environmental context in which these behaviors occurred. The terms detection of risky behaviors, driving risk probability, driver scoring, and driver profiling, that are used interchangeably in literature, are clearly distinguished and mathematically defined.
- 2) An ensemble supervised machine learning algorithm based on randomized trees is selected and customized to reflect the predicted driving risk probability. Driving behaviors and environmental contextual data are both utilized as inputs to the selected model. The model is

proven to provide an acceptable compromise between bias and variance as shown in the results section.

- 3) Risk prediction model is trained and tested using an unprecedented amount of real driving data using SHRP2 NDS. This enhances reliability and practicability of proposed system.

The paper is structured as follows. In section II, we provide a background information on the driving dataset used in this work. Similar driver profiling frameworks presented in literature are also discussed. Section III provides a detailed description of the envisioned CEDP system with a focus on risk prediction module. In section IV, the adopted pre-processing, error analysis and model selection processes for risk prediction problem are described. In section V, results are presented and discussed. Conclusions and future work are presented in section VI.

II. BACKGROUND AND RELATED WORK

A. Driving Behavior Profiling

In literature, the term driver behavior profiling has been used to describe different behavioral characterization processes, which may cause confusion. Some of the literature work have used the driver profiling term interchangeably with behavior classification or detection. For instance, authors in [8] utilized variations of Recurrent Neural Network (RNN) models to detect seven distinct types of behaviors using smartphone sensors. In similar work, authors in [9] used static supervised machine learning techniques such as Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) to detect certain driving maneuvers. Likewise, in [10], authors proposed a sequence modelling HMM-based classifier to classify aggressive and normal driving maneuvers in both forward and lateral directions. Classification was based on smartphone sensory data with a classification accuracy of 95 %. Similar work is presented in [11]. Although all of the aforementioned work appeared in the context of driver profiling, rather to be more precise it should be under the driving behavior classification/detection umbrella.

Driver profiling is the process of augmenting different driving behaviors, over several driving trips, into a scoring function to measure a driver's overall competence level. Notable research in this context is the work presented in [2]. Authors in this paper have made a clear distinction between behavior detection and driver profiling. Using a fuzzy logic algorithm hosted by a smartphone application, four unique driving events i.e., harsh braking, aggressive acceleration, speeding, and aggressive steering were accurately detected by acquiring a smartphone's accelerometer, gravity, magnetic, and GPS data. A scoring function was then introduced to reflect the overall driving trip score given the detected behaviors. Despite the proposals and findings of the paper, the scoring function was very primitive, since it did not reflect the statistical correlation between actual risk and detected behaviors. Moreover, it did not show how to find an overall driving profile as a function of many trips. In other words, it did not elaborate how the

individual trips' scores will be used towards building a driver's profile.

Recently, a more rigorous work has been presented in [12]. Authors in this paper have presented a data-driven scoring system using SHRP2 NDS. The behavioral information of a very large number of driving events, as well as total driving time were used to predict driving risk using supervised machine learning algorithms. A driving score was then formulated as an additive inverse of the predicted driving risk probability. Despite the promising performance results in [12], the joint effect of driving behaviors and the environment in which such behaviors occurred was ignored.

To the best of our knowledge, no work in the literature has considered a complete driver behavior profiling system that considers the sub-processes of behavior detection, risk prediction, driver's behavior scoring and profiling, and with the consideration of driving environment. In this paper, an envisioned data-driven profiling system is introduced and discussed. Specifically we targeted the problem of driving risk prediction utilizing behavioral and environmental data of a large scale NDS. The development of the risk prediction model is based on an error analysis of different supervised machine learning models to achieve the best bias-variance trade-off.

B. Dataset

We utilized the SHRP2 NDS dataset [13], which is the largest NDS to date. Row data contains the detailed information of more than 29,000 driving events. Detailed information includes behaviors that are apparent within seconds before risky events or during captured normal driving episodes. Behaviors in the context of this work are different than the *in-vehicle* distractions. They are *vehicle-kinematic* observations that can be noticed from outside the vehicle such as: aggressive driving, speeding, etc. In addition to driving behaviors, SHRP2 NDS has the environmental contextual information at which these behaviors happened. Environmental information can be categorized into three types: static (e.g., road curvature), quasi-static (e.g., road lighting), and dynamic (e.g., traffic density).

A driving event in SHRP2 is one of the following types [7]: a crash, a near-crash, a non-subject conflict, and a balanced baseline.

In this work, behaviors and the three aforementioned environmental categories are used as predictors to risk, quantified herein in terms of crash and near crash events.

III. PROPOSED DRIVER PROFILING FRAMEWORK

In this section, the proposed cloud-based environment-aware driver profiling framework is discussed. The discussion will cover the details of the complete driver profiling system, from the in-vehicle data acquisition to the cloud-based profiling. Unlike other profiling systems, the proposed system's risk prediction is motivated by statistically significant results as will be shown in section V. Figure 1 depicts the framework block diagram. The proposed framework consists of two main modules, in-vehicle and cloud.

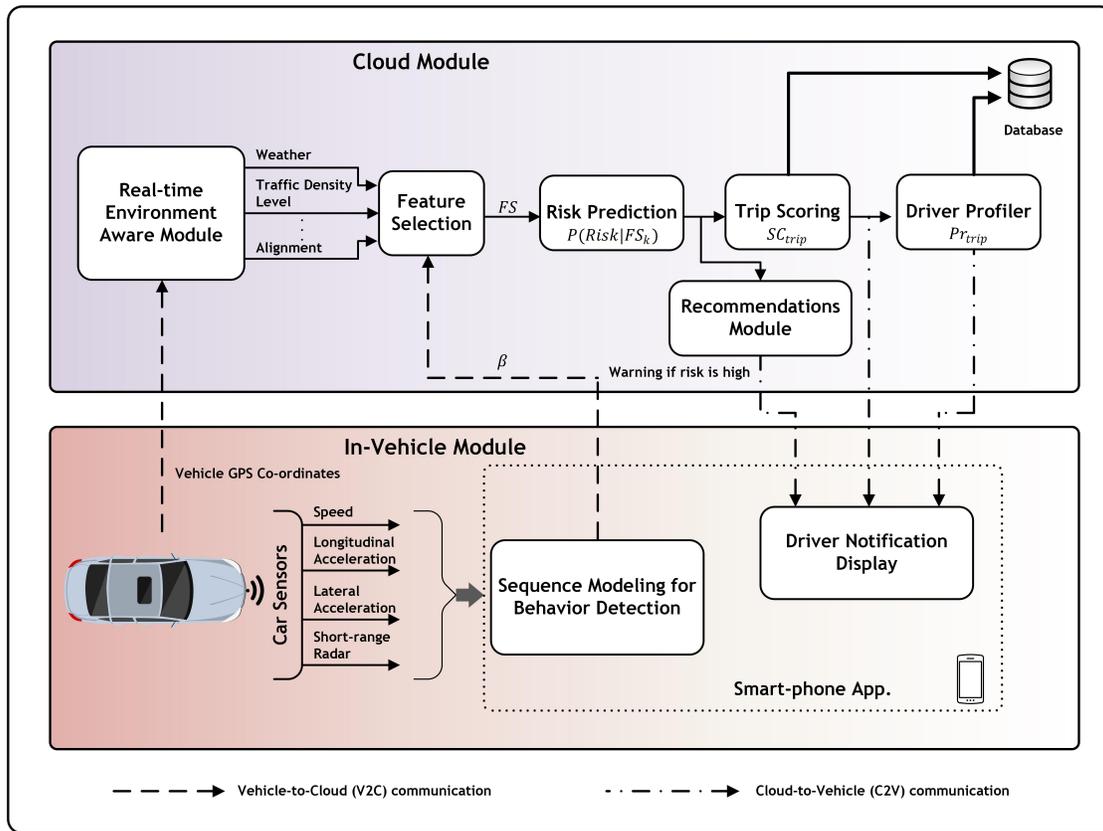


Fig. 1: Proposed Cloud-Based Environment-Aware Driver Profiling Framework.

A. In-Vehicle Module

The in-vehicle module contains data collection, pre-processing and modeling processes that occur inside the vehicle. In this module, collected data can be divided into two types:

- 1) *Type 1*: Data that reflects the longitudinal and lateral behavior of the vehicle. This data is collected through the vehicle's Controller Area Network (CAN) bus by utilizing an On-Board Diagnostics II (OBDII) unit or through smartphone sensors.
- 2) *Type 2*: Data that reflects the relative position of the subject vehicle to the surrounding vehicles and provides driving context awareness. This is gathered using short range radar (SRR) sensors.

Once data is collected, it is sent from OBDII and radar interfaces to the subject driver's smartphone through a bluetooth link. In the smartphone, a hosted application acquires these time-series data every T_s seconds and applies sequence modeling for behavior detection (e.g., HMM-based Modeling). The application outputs a detected behavior (β) and sends it to the cloud, along with the subject vehicle's GPS co-ordinates.

B. Cloud Module

Inside the cloud, the vehicle's GPS co-ordinates are inputted to a real-time environment aware module to find the environmental attributes which the vehicle is subjected to. This

module returns the corresponding attributes: weather condition (W), traffic density level (TD), road lighting conditions (L), traffic control (TC), road flow (RF), and road alignment (A). The returned environmental attributes vector env , along with the driver's detected behavior β form the initial feature vector F :

$$F = [\beta, env] \quad (1)$$

Based on a pre-trained risk prediction model, some of the irrelevant initial environmental attributes are discarded to enhance the prediction model's performance (e.g., reduce over-fitting). The selected attributes form the feature vector FS , where $FS \subset F$ and is expressed as:

$$FS = [\beta, env'] \quad (2)$$

The pre-trained risk prediction model uses FS to predict the driving risk probability $P(Risk|FS)_k$, where the subscript k is an integer that represents an event index. The driving risk probability is expressed mathematically as:

$$P(Risk|FS)_k = P(C|FS)_k + P(NC|FS)_k \quad (3)$$

where $P(C|FS)_k$ and $P(NC|FS)_k$ are, respectively, the conditional probabilities of crash and near-crash events given the feature vector FS at event k . Once the risk probability is calculated, a warning is issued to the subject driver if

TABLE I: Summary of Notations

Notation	Description
F	Initial feature vector
FS	Selected features vector
$P(Risk FS)_k$	Risk probability of event k given FS
β	A detected behavior
env	A vector with initial environmental attributes
env'	A vector with selected environmental attributes
$N_{\beta,env'}$	Number of captured events containing β in env' , after the first warning about β
$\bar{N}_{\beta,env'}$	Total number of captured events in env' , after the first warning about β
$c_{\beta}^{env'}$	Compliance with a warning about β captured in env'
C_{trip}	Driver's overall compliance per in $trip$
N_{tot}	Total number of captured risky events per trip, after the first warning of each
N_{env}	Total number of driving environments in a certain trip
N_{β}	Number of different types of detected risky behaviors in a selected environment env' .
P_{trip}	Average risk probability in $trip$
Sc_{trip}	Driver's score in $trip$
Pr_{trip}	Driver's risk profile after $trip$.
N	Total number of captured events in a trip
ξ	Weight of EMWA filter

$P(Risk|FS)_k > p_{th}$. The driver's compliance to the warnings along with a weighted sum of the aggregated risk probabilities over a certain trip are both used to compute the final trip score Sc_{trip} .

A driver's compliance to a warning about behavior β in env' is measured as a function of the number of risky events in env' (i.e., $P(Risk|FS)_k > p_{th}$) that involve β after receiving the first warning about it. Let $c_{\beta}^{env'}$ represents the driver's compliance to the warning about the risky behavior β in driving environment env' . $c_{\beta}^{env'}$ may be computed using the following utility function:

$$c_{\beta}^{env'} = 1 - \left(\frac{N_{\beta,env'}}{\bar{N}_{\beta,env'}} \right) \quad (4)$$

where $N_{\beta,env'}$ and $\bar{N}_{\beta,env'}$ are, respectively, the number of events that contain the behavior β in environment env' , and the total number of events captured in env' , after the first warning about β .

The driver's overall compliance per trip C_{trip} can be measured as the average compliance ratio for various β in different env' . This can be mathematically expressed as:

$$C_{trip} = \left(\frac{1}{N_{tot} + 1} \right) \cdot \sum_{env'=1}^{N_{env}} \sum_{\beta=1}^{N_{\beta}} c_{\beta}^{env'} \quad (5)$$

where N_{tot} is the total number of risky behaviors after the first warning of each of them, N_{env} is the total number of driving environments in a certain trip, and N_{β} is the total number of detected risky behaviors in a certain environment env' . $C_{trip} \in [0, 1]$, where a zero value indicates that the

driver was non compliant with the issued warnings, while a value of one reflects a compliance with all warnings.

The trip score is then computed as a function of the trip weighted sum of the aggregated risk probabilities P_{trip} , and the driver's compliance value per trip C_{trip} :

$$Sc_{trip} = \mathcal{F}(C_{trip}, P_{trip}) \quad (6)$$

P_{trip} can be simply calculated as the trip average risk probability, which is denoted by the following formula:

$$P_{trip} = \frac{1}{N} \sum_{k=1}^N P(Risk|FS)_k \quad (7)$$

where N is the total number of captured events in a trip. Considering a normalized $Sc_{trip} \in [0, 1]$:

$$Sc_{trip} = \gamma \cdot C_{trip} + \alpha \cdot (1 - P_{trip}) \quad (8)$$

where

$$\gamma + \alpha = 1 \quad (9)$$

The values of γ and α determine how much weight is given to C_{trip} and P_{trip} . For instance, if $\alpha = 1$, the overall trip score will be determined solely based on the value of P_{trip} (i.e., $\gamma = 0$).

Finally, a subject driver's profile after a certain trip (Pr_{trip}) can be computed using an exponentially moving weighted average (EMWA) filter applied on various trip scores to assign exponentially increasing weights for recent trips. This is expressed as:

$$Pr_{trip} = \begin{cases} Sc_1, & \text{if } trip = 1 \\ \xi \cdot Sc_{trip} + (1 - \xi) \cdot Pr_{trip-1}, & \text{if } trip > 1 \end{cases} \quad (10)$$

where the value of ξ determines the number of trips which the filter will use to calculate Pr_{trip} .

An important feature of the presented framework is the prediction of driving risk probabilities given the behavioral and environmental attributes. Non-accurate values of these probabilities can result in missed or false warnings as well as unreliable driving scores. The rest of the paper contains the necessary steps for the development of the driving risk prediction model and its performance assessment.

IV. DATA PRE-PROCESSING AND MODEL SELECTION

A. Data Pre-processing

Row data contains the information of $\sim 29,000$ driving events, each with a certain severity level. In the original dataset, event severity levels are exclusively contained in the following set: $Severity = \{Crash, Near-Crash, Crash-Relevant, Non-Subject Conflict, Balanced Baseline\}$. Each event is represented by a vector that contains the captured driving behavior of the subject driver prior to a risky event (or during a normal driving event), and the environmental context in which these behaviors happened.

TABLE II: Summary of Environmental Conditions

Traffic Flow	Traffic Density	Traffic Control	Weather Conditions	Lighting Conditions	Road Alignment
Divided	Stable	Yes	No Adverse Conditions	Dark	Straight
Not Divided	Stable With Flow Restrictions	No	Foggy	Lighted	Curved
No Lanes	Unstable	-	Rainy	-	-
-	-	-	Snowy	-	-

1) *Data Merging*: In this work, *Crash* and *Near-Crash* severity levels are put under the common severity level of *Risky*, whereas *Non-Subject Conflict* and *Balanced Baseline* events are used to represent the *Normal* level. Under each environmental category, similar features are merged to increase their importance in order to enhance the prediction model performance (e.g., under road alignment category, curved to the right and curved to the left features are considered the same). Similarly, 12 main behaviors are identified from the original behaviors as it has been discussed in [12]. Identified environmental features are shown in table II.

2) *Data Encoding*: After data merging, the behavioral and environmental categorical variables are encoded to integers. To calculate risk probability, events with same behavioral and environmental features are combined and the corresponding risk probability for each is calculated. To represent data in a meaningful way for machine learning algorithmS, the *one-hot encoding* technique is utilized.

B. Model Selection

After data is encoded, it is divided to training and development sets according to the ratio 75% and 25%, respectively. Using accuracy as a performance metric, an error analysis for a simple multiple linear classification model indicated a high bias (i.e., low training set accuracy). More complex structured SVM-based models, on the other hand, were able to model training data accurately, but were not capable of generalizing on the development set (i.e., high variance). To achieve a good bias-variance trade-off, a customized random forest model was selected. In random forests, multiple decision trees are built, each from a sample of the training set. The best split in each tree is based on a random subset of the input features rather than the whole feature set. The average performance of the various trees is then used to reflect the forest performance. Although this approach theoretically causes a slight degradation in the training set performance, it reduces over-fitting due to the averaging process. In this work, a customized random forest model resulted in the best bias-variance performance in both classification and regression contexts.

The adopted hyper-parameters of the selected model are shown in table III, where M represents the number of all behavioral and environmental features, and MSE is the mean square error.

V. RESULTS AND DISCUSSION

This section presents the performance results of the Random Forests risk prediction model presented in section IV. The model was implemented in Spyder (Python 3.6) integrated

TABLE III: Hyper-parameters of RF Model

Hyper-parameter	Classification	Regression
<i>Number of Trees</i>	100	100
<i>Split Criterion</i>	Entropy	<i>MSE</i>
<i>Max No. of Features per Tree</i>	\sqrt{M}	M

development environment (IDE) using the Scikit-Learn Library for Machine Learning and Data Mining. Results in both classification and regression contexts are discussed. Reported results are those obtained from the customized RF model after trying different random seeds. They represent the best obtained results.

A. Classification

The driving risk prediction problem is formulated as a binary classification problem where:

$$Risk = \begin{cases} 1, & \text{if } P(Risk|FS)_k > 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

Five performance measures, accuracy, precision, recall, specificity, and the area under the ROC curve (AUC), are utilized to reflect the model performance. Figure 2 depicts the AUC performance of the RF vs. SVM classifiers. The figure shows that the proposed RF classifier possesses a high True Positive Rate (TPR) of $\sim 93\%$ given an insignificant False Positive Rate (FPR) of $\sim 5\%$, with an average AUC value of 0.98. The obtained TPR and FPR ratios should ensure that the number of inaccurate warnings to a subject driver during a certain trip are minimized.

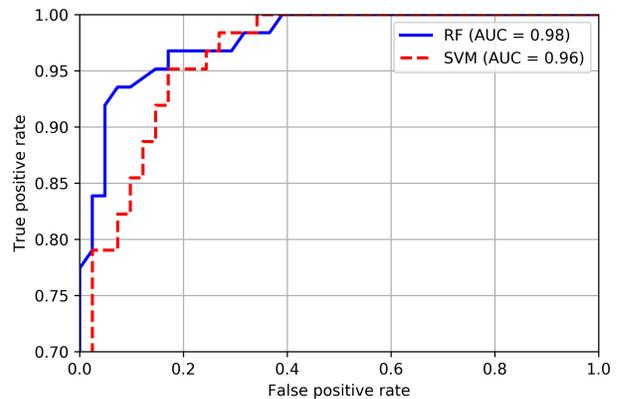


Fig. 2: Receiver Operating Characteristic (ROC) curve for selected RF classifier.

Table IV summarizes the other results. It can be seen from the training set performance that the model has a zero bias, indicating that all the training samples are correctly classified. Nevertheless, the development set performance shows an average of 6.4 % degradation, which indicates a slight over-fitting. Although this bias-variance combination was the best achieved, a thorough troubleshooting of the over-fitting causes is still needed.

TABLE IV: Classification Performance Results

Performance Measure	Training Set (%)	Development Set (%)
Accuracy	100.0	93.2
Precision	100.0	95.08
Recall	100.0	93.5
specificity	100.0	92.7

B. Regression

While the classification performance results are pivotal to avoid false or missed warnings, the prediction of the soft risk probabilities (i.e., regression results) is crucial in calculating the overall driving risk score. In this work, Mean Square Error (MSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2) are used to reflect the model performance from a regression perspective. MSE training results indicate that the model is unbiased, with $MSE < 1\%$. However, the difference between the MSE of the training and development sets as well as the discrepancies between their MAE values show that the model is slightly over-fitted.

TABLE V: Regression Performance Results

Performance Measure	Training Set	Development Set
MSE (%)	0.654	2.6
MAE (%)	5.7	12.1
Adjusted R^2	0.948	0.776

Nevertheless, from a practical standpoint, MAE results of the development set show that the average absolute deviation from the true risk probabilities is 12.1%, which is still an acceptable performance and would not affect the overall profile score. Another adopted statistical measure to check the model robustness is the adjusted R^2 criterion. It reflects how well the model interprets the changes of data around its average value. Although there is a noticeable difference between training and development sets' R^2 performances, both R^2 values are vigorous, since they possess values greater than 0.5 in both cases.

VI. CONCLUSION

In this paper, an envisioned novel driver profiling framework was thoroughly presented and discussed. The framework contains data processing on both vehicle and cloud levels. The paper specifically addressed the risk prediction problem by utilizing the behavioral and environmental contextual information of 29,000 driving events, using the SHRP2 NDS. By analyzing the error of different models, a customized randomized trees model appears to give the best bias-variance

trade-off in classification and regression contexts. Results confirm that behavioral and environmental data are together good predictors to driving risk, which is measured in this paper in terms of crash, near-crash and crash-relevant events. In future work, a thorough analysis of development set error sources is warranted. Moreover, testing the developed model on a new dataset would ensure the robustness of the model.

ACKNOWLEDGMENT

The findings and conclusions of this paper are those of the authors and do not necessarily represent the views of the Virginia Tech Transportation Institute, SHRP 2, the Transportation Research Board, or the National Academy of Sciences.

This research is supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) under grant number: STPGP 479248.

REFERENCES

- [1] "WHO | Global status report on road safety 2015." [Online]. Available: http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/
- [2] G. Castignani, T. Derrmann, R. Frank, and T. Engel, "Driver Behavior Profiling Using Smartphones: A Low-Cost Platform for Driver Monitoring," *IEEE Intelligent Transportation Systems Magazine*, vol. 7, no. 1, pp. 91–102, 2015.
- [3] D. I. Tselentis, G. Yannis, and E. I. Vlahogianni, "Innovative Insurance Schemes: Pay as/how You Drive," *Transportation Research Procedia*, vol. 14, pp. 362–371, Jan. 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2352146516300898>
- [4] M. J. Davidson and J. A. O. III, "Systems and methods for utilizing telematics data to improve fleet management operations," US Patent US8416067B2, Apr., 2013. [Online]. Available: <https://patents.google.com/patent/US8416067B2/en>
- [5] K. Takeda, C. Miyajima, T. Suzuki, P. Angkititakul, K. Kurumida, Y. Kuroyanagi, H. Ishikawa, R. Terashima, T. Wakita, M. Oikawa, and Y. Komada, "Self-Coaching System Based on Recorded Driving Data: Learning From One's Experiences," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1821–1831, Dec. 2012.
- [6] P. Handel, I. Skog, J. Wahlstrom, F. Bonawiede, R. Welch, J. Ohlsson, and M. Ohlsson, "Insurance Telematics: Opportunities and Challenges with the Smartphone Solution," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 4, pp. 57–70, 2014.
- [7] "SHRP2 NDS Data Access." [Online]. Available: <https://insight.shrp2nds.us/home>
- [8] E. Carvalho, B. V. Ferreira, J. Ferreira, C. d. Souza, H. V. Carvalho, Y. Suhara, A. S. Pentland, and G. Pessin, "Exploiting the use of recurrent neural networks for driver behavior profiling," in *2017 International Joint Conference on Neural Networks (IJCNN)*, May 2017, pp. 3016–3021.
- [9] J. F. Júnior, E. Carvalho, B. V. Ferreira, C. d. Souza, Y. Suhara, A. Pentland, and G. Pessin, "Driver behavior profiling: An investigation with different smartphone sensors and machine learning," *PLOS ONE*, vol. 12, no. 4, p. e0174959, Apr. 2017. [Online]. Available: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0174959>
- [10] S. Daptardar, V. Lakshminarayanan, S. Reddy, S. Nair, S. Sahoo, and P. Sinha, "Hidden Markov Model based driving event detection and driver profiling from mobile inertial sensor data," in *2015 IEEE SENSORS*, Nov. 2015, pp. 1–4.
- [11] A. Abdelrahman, N. Abu-Ali, and H. S. Hassanein, "Driver Behavior Classification in Crash and Near-Crash Events Using 100-CAR Naturalistic Data Set," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Dec. 2017, pp. 1–6.
- [12] A. Abdelrahman, H. S. Hassanein, and N. Abu-Ali, "Data-driven Robust Scoring Approach for Driver Profiling Applications," in *GLOBECOM 2018 - 2018 IEEE Global Communications Conference*, Dec. 2018.
- [13] J. M. Hankey, M. A. Perez, and J. A. McClafferty, "Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets," Apr. 2016. [Online]. Available: <https://vtechworks.lib.vt.edu/handle/10919/70850>