Machine Learning-Based Approach For Detecting Driver Behavior Using Smartphone Sensors

Mohammad Abbadi, Ahmad Abadleh, Saqer Alja'afreh, Zaid Halhouli

Abstract: This paper aims at combining machine learning techniques with Smartphone sensors (i.e. accelerometer sensor) to develop a smart model capable of classifying vehicle driving style into (Excellent, good or weak) categories. In this paper, we use several machine learning algorithms (Neural network, KNN, Naïve Bayes and Random forest tree) to train and test data extracted from Smartphone sensors. The results indicate the possibility to exploit Smartphone sensor readings in the design of a reliable model capable of identifying the driving style based on accelerometer readings. All examined machine learning algorithms maintain high accuracy in classifying the vehicle's driving class; however, the neural network classifiers achieve the highest accuracy ratio reaches (99.9%).

Index Terms: Smartphones, accelerometer sensor, car driving, Machine learning.

1. INTRODUCTION

Car driving is a key part of daily life for individuals to meet their needs or go to work or even take a walk. Drivers differ in their driving styles as well as their commitment to traffic rules. Good car driving basically requires certain behavior and driving skills that may not be linked to traffic rules. In many cases roads are not entirely suitable for driving; there may be bumps, potholes, or rough roads. This requires the driver to act in a way that can be reflected positively on avoiding problems or accidents that may result from the presence of these obstacles. However, the only way to master these skills is by experience for long driving years. Cities are flooded with vehicles and roads are constantly changing because of the tremendous increase in vehicles and active traffic. This puts an extra burden on drivers to drive properly without endangering their lives and the lives of others as well as the vehicle [1], [2]. A direct impact of this context can reduce traffic safety since city drivers become more aggressive in driving. Therefore, it is important to develop a judgmental tool on a driving style based on their driving behavior, which can be categorized by using several driving parameters. However, finding a mechanism that defines a good driving style based on these parameters may not be an easy task. Therefore, an urgent need has been raised to develop a method in which appropriate behaviors can be identified to ensure good daily and personal driving [3], [4], [5], [6], [7]. These attempts to find such method mostly are descriptive; in other words, qualitative measures such as attitudes and views of drivers about good driving. While very few studies involve virtual or smart driving simulator to collect empirical driving data from human drivers and to understand human driving behavior [8] or categorize driving style by combining expert's evaluation with machine learning techniques based on data generated by smartphone accelerometer sensor [9]. Literature provides several models for virtual traffic flow and decision making by drivers in exceptional circumstances such as emergency or natural disaster [10]. Unfortunately, very few studies have attempted to develop experience in driving cars in the form of good driving models at the individual level and regular daily driving whether for urban or suburban traffic [10], [7], [11], [12].

- *Mohammad Abbadi, Department of Computer Science Faculty of IT, Mutah University.*
- *Ahmad Abadleh, Department of Computer Science Faculty of IT, Mutah University.*
- *Saqer Alja'afreh, Dept of Electrical Engineering Faculty of Engineering, Mutah University.*
- *Zaid Halhouli, Department of Computer and IT Faculty of Science, Tafila Technical University.*

2 RELATED WORK

Recently, there has been increasing interest in controlling driver behavior, driving style and traffic settings to avoid accidents and save lives. Various methods have emerged in the literature, such as special composite cameras, sensors, advanced driver assistance systems (ADAS), etc. to control roads, traffic settings and driver behavior. In [13], they emphasized the use of smartphones sensors to monitor driving behavior as one of the most cost-effective methods. In their study, they developed and implemented a system that uses a smartphone accelerometer sensor to detect sudden changes in acceleration; however, the system identifies insecure / sharp turns using the gyroscope sensor in an operative manner. The system categorizes the driver as an aggressive driver or vice versa based on the observed driving pattern.

To categorize drivers according to their driving style, machine learning algorithms and smartphone data combined in [8]. They used separate Haar wavelet transformation to remove noise effects. After that, they obtain a separate wavelet transformation with four levels of data from the smartphone sensors, which include low to high frequencies respectively. The vector includes the features obtained for each rough signal variation maneuver as well as the variation of wavelet transformation components. They used KNN to select the features from these vectors and SVM, RBF and MLP neural networks to recognize braking maneuvers and dangerous speeds from safe maneuvers as well as dangerous ones. The features obtained by discrete wavelet transformation were very useful for maneuvers assessment. In [9], they offered innovative technologies to improve the accuracy and usability of smartphone sensors. Machine learning algorithms were employed to detect changes in the relative direction of the smartphone resulting from human interactions. Then, they developed graph-aware alignment algorithm to increase alignment accuracy and track the linear acceleration of the vehicle to address the problems of over / under acceleration. The results lead to develop a smartphone application called XSense that uses new technologies to improve overall accuracy in driving analytics. XSense improved 75% accuracy compared to inertial sensors that are well-tuned in the conventional approach. Another system called Nericel designed by Mohan et al [13], is a Windows Mobile Smartphone application for monitoring road and traffic conditions. The accelerometer is used to detect bumps and break events. It also uses a microphone to detect GPS / (GSM) for car localization and speed. Driving events such as braking, bumps and potholes are identified by comparing a set of predetermined thresholds with sudden variations in the accelerometer data or average during a sliding window of N seconds. Results include: 4.4% FN, 22.2% FP for urgent events, 23% FN, and 5% FP for bump detection at low speed (<25 km / Hour); and 0% FN and FP to detect the honking on an exposed vehicle. Dai and colleagues [14] suggested an Android app that aims to provide a real-time detection and alerting to dangerous driving events typically associated with DUI instruction. The application uses smartphone accelerometer and direction sensors (angular, tilt, and fascia angles) to detect abnormal curvature and speed maintenance problems, which are two main categories of drunk driving behaviors. An abnormal curvature event is detected if the difference between the maximum and minimum lateral acceleration values over a 5 second period exceeds the experimental threshold. Maintenance problems are detected if the longitudinal acceleration exceeds the static or negative-positive thresholds at any time. Similarly, any MLA does not work for event detection. The experimental results include: 0% FN, 0.49% FP for abnormal curved movements; 0% FN, and 2.90% FP for speed control problems. Fazeen and colleagues [15] used an accelerometer of Android smartphones and GPS to determine car conditions (speed and shifting), driving styles (acceleration / deceleration and change of pathways) and road conditions (smooth, uneven, rough, bumpy or punctured). The events are mainly detected by calculating the time, difference, and slope between successive accelerometer readings on specific axes and comparing them with static / dynamic experimental thresholds. For instance, work states that safe acceleration and deceleration never reaches a g-force more than ± 0.3g on the y-axis. No MLA is used to detect the event. The experimental results include the accuracy of the classification of the anomalies as the following: 81.5% for bumps, 72.2% for potholes, 75% for the rough road, 91.5% for smooth roads, 89.4% for uneven roads.

3 METHOD

We have designed this paper under a supervised learning process where the extracted features lead to categorizing car driving profiles as multi-label behaviors of excellent, good or weak. This paper aims at determining the best combination of accelerometer sensor data (i.e. sensor axis(es)) and a learning algorithm to detect individual driving profiles. Fig.1 defines our model and the evaluation metric.

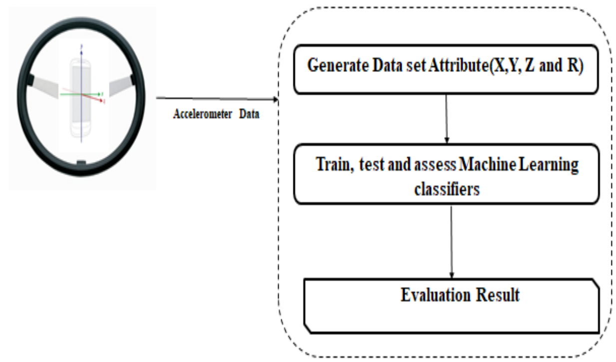

Fig. 1. *. Study Model*

Fig.1 shows a view of our model architecture. In the first step of the model, smartphone accelerometer sensor raw data is sampled while driving. Sampled sensor data are then stored in the smartphone file system. In the second step, stored sensor data files are retrieved from the smartphone and used as input to generate our data sets attributes. In the third step, data sets attributes are entered to train, test and assess machine learning classifier performances. As shown in Fig. 1, data collection is carried out by using the smartphone, whereas the next three steps are executed on a computer using Orange software, which is an open-source software package released under GPL [15].

4 Evaluation

To evaluate our model, we compare the performance of several machine learning classifiers: Artificial Neural Networks (ANN), Random Forest (RF), and Bayesian Network (BN) [2], [16], [17], [18]. We have chosen these classifiers because of their extensive use in literature as well as their ability to process diverse data [18]. The metric we use to evaluate machine learning classifiers performance for each driving category is the accuracy based on the confusion matrix as shown in equation 1, which ranges from 0.0 (worst) to 1.0 (best) [17]. Therefore, the closer evaluation of classifier accuracy is to 1.0, the better it is at perceiving a driving profile. Fig. 2 shows our evaluation metric to using Orange software to ensure the validly of our excellent, good or weak driving model.

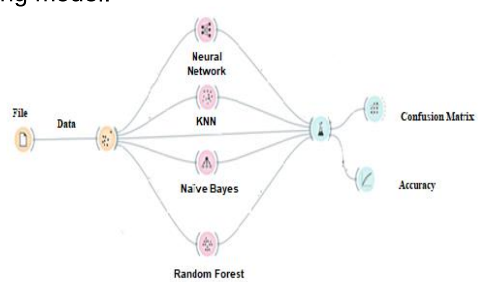

Fig. 2. *Evaluation Metric*

Accuracy=(TP+TN)/(P+N) ………………. (1)

Where TP denotes True Positive, TN denotes True Negative, P denotes positive cases and N denotes Negative cases.

5 DATA COLLECTION

We conducted a real-world experiment in order to collect accelerometer sensor data for driving category. In our experiment, a specially designed application developed to record the smartphone accelerometer sensor data while a driver is driving for a particular time. The data were recorded to determine the rating of each driver according to the intended classification (Excellent, Medium, Weak) by collecting data for different drivers at the same time and the same way at three different speeds of 40, 70 and 100. We performed the experiment using more than one type of car as well as different models. Four volunteer drivers drove the cars for 15 minutes each at different speeds. During the driver's trip, the smartphone was horizontally fixed and installed on the surface of the car's tabloid to ensure accuracy in reading the sensing data. The driving categories we modelled in this experiment based on accelerometer sensor data. Our aim was to define a set of driving attributes that represents excellent, good or weak driving style. The data set contains four feature attributes: sensor axis(es), x, y and z, acceleration, R and driving class (Excellent, Medium, Weak).

6 RESULT

We performed all experiments for collecting data and combining machine learning classifiers and accelerometer sensor data to build our dataset described in the previous section. Table 1 shows a sample of our data set representing extracted features and classes.

TABLE 1

SAMPLE OF OUR DATA SET

ID	Time(s)	X(m/s2)	Y(m/s2)	Z(m/s2)	R(m/s2)	Class
1	386960.8	1.85	1.72	9.35	9.68	excellent
2	386961.1	1.82	1.51	9.50	9.79	excellent
3	386961.5	1.87	1.61	9.43	9.74	excellent
4	386961.7	1.97	1.59	9.47	9.80	excellent
5	386962.1	1.81	1.47	9.43	9.72	excellent
6	386962.5	1.51	0.80	9.60	9.75	excellent
7	386963.0	0.98	0.44	10.12	10.18	excellent
8	46.78	2.13	1.84	5.11	5.83	good
9	46.79	1.93	1.55	4.86	5.45	good
10	46.79	1.71	1.24	4.62	5.08	good
11	46.80	1.38	0.98	4.41	4.72	good
12	386989.16	1.52	1.38	9.75	9.97	weak
13	386989.16	1.44	1.76	8.61	8.91	weak
14	386989.19	1.65	1.99	7.73	8.15	weak
15	386989.22	1.62	2.14	7.32	7.79	weak

Afterwards, we trained, tested, and assessed every machine learning classifier with its default configuration using the sampling type of stratified 10 -fold cross-validation in order to minimize overfitting. Finally, we obtained the confusion matrix for these executions, calculated the accuracy for each classifier, and ranked the best performing classifier. This section presents the result of our evaluation metric. We trained and tested (22954) cases using the orange software and based on the default parameters of the chosen machine learning classifiers. Stratified 10 -fold cross-validation was used for sampling technique. Fig. 3 shows the resulted confusion matrix for each classifier.

Neural Network

Actual \ Predicted	excellent	good	weak	∑
excellent	2425	106	111	2642
good	89	18007	72	18168
weak	56	44	2044	2144
∑	2570	18157	2227	22954

Naïve Bayes

Actual \ Predicted	excellent	good	weak	∑
excellent	1954	462	226	2642
good	876	16791	501	18168
weak	13	372	1759	2144
∑	2843	17625	2486	22954

Random Forest

Actual \ Predicted	excellent	good	weak	∑
excellent	2417	124	101	2642
good	82	18040	46	18168
weak	59	62	2023	2144
∑	2558	18226	2170	22954

KNN

Actual \ Predicted	excellent	good	weak	∑
excellent	2482	70	90	2642
good	115	18000	53	18168
weak	98	36	2010	2144
∑	2695	18106	2153	22954

Fig. 3. *Confusion Matrix for chosen classifiers.*

Based on the above result and equation 1, we can conclude that our chosen classifiers achieved the flowing accuracy level as shown in table 2:

TABLE 2

ACCURACY LEVEL FOR MACHINE LEARNING CLASSIFIERS

Classifier	Accuracy (%)
Neural Network	99.9%
KNN	99.5%
Naïve Bayes	96.2%
Random Forest	99.8%

All classifiers maintain a high level of accuracy; however, the neural network is the best performing classifiers as it achieved (99.9%) of accuracy level.

7 CONCLUSIONS

In this paper, we modelled an approach for combining smartphones sensor data with machine learning techniques to identify car driving style wither excellent, good or weak. Real experiments were to be carried out to record smartphone readings while driving different types of cars at different speeds. Collected readings designed in the form of a training and testing datasets. Several machine learning techniques applied to evaluate our approach. The results showed the feasibility and usability of the approach to predict the driving class through high accuracy ratios; however, the highest accuracy rate was in favor of neural network techniques with (99.9%).After proving the possibility of determining the style of driving with high accuracy; as future work, this work can be pursued by attempting to extract events performed by the driver to maintain good or weak driving considering natural obstacles such as bumps, the roughness of the road and traffic. Additional smartphone sensor might need to be involved for further enhancement.

REFERENCESS

[1] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software: an update,” ACM SIGKDD Explor. Newsl., vol. 11, no. 1, pp. 10–18, 2009.
[2] T. K. Ho, “Random decision forests,” in Proceedings of 3rd international conference on document analysis and recognition, 1995, vol. 1, pp. 278–282.
[3] Z. Constantinescu, C. Marinoiu, and M. Vladoiu, “Driving style analysis using data mining techniques,” Int. J. Comput. Commun. Control, vol. 5, no. 5, pp. 654–663, 2010.
[4] A. A. Tecimer, Z. C. TAYŞİ, A. L. İ. G. YAVUZ, and M. E. K. YAVUZ, “Assessment of vehicular transportation quality via smartphones,” Turkish J. Electr. Eng. Comput. Sci., vol. 23, no. Sup. 1, pp. 2161–2170, 2015.
[5] R. F. Tinder, “Relativistic flight mechanics and space travel,” Synth. Lect. Eng., vol. 1, no. 1, pp. 1–140, 2006.
[6] W. Rindler, Essential relativity: special, general, and cosmological. Springer Science & Business Media, 2012.
[7] R. Caruana and A. Niculescu-Mizil, “An empirical comparison of supervised learning algorithms,” in Proceedings of the 23rd international conference on Machine learning, 2006, pp. 161–168.
[8] R. Lotfi and M. Ghatee, “Smartphone based Driving Style Classification Using Features Made by Discrete Wavelet Transform,” arXiv Prepr. arXiv1803.06213, 2018.
[9] L. Kang and S. Banerjee, “Practical driving analytics with smartphone sensors,” in 2017 IEEE Vehicular Networking Conference (VNC), 2017, pp. 303–310.
[10] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, p. 27, 2011.
[11] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” Proc. Natl. Acad. Sci., vol. 79, no. 8, pp. 2554–2558, 1982.
[12] E. Fix and J. L. Hodges, “Discriminatory analysis. Nonparametric discrimination: consistency properties,” Int. Stat. Rev. Int. Stat., vol. 57, no. 3, pp. 238–247, 1989.
[13] R. Chhabra, S. Verma, and C. R. Krishna, “Detecting Aggressive Driving Behavior using Mobile Smartphone,” in Proceedings of 2nd International Conference on Communication, Computing and Networking, 2019, pp. 513–521.
[14] J. Dai, J. Teng, X. Bai, Z. Shen, and D. Xuan, “Mobile phone based drunk driving detection,” in 2010 4th International Conference on Pervasive Computing Technologies for Healthcare, 2010, pp. 1–8.
[15] M. Fazeen, B. Gozick, R. Dantu, M. Bhukhiya, and M. C. González, “Safe driving using mobile phones,” IEEE Trans. Intell. Transp. Syst., vol. 13, no. 3, pp. 1462–1468, 2012.
[16] E. M. Kleinberg, “On the algorithmic implementation of stochastic discrimination,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 5, pp. 473–490, 2000.
[17] T. Fawcett, “An introduction to ROC analysis,” Pattern Recognit. Lett., vol. 27, no. 8, pp. 861–874, 2006.
[18] D. Beerbaum and J. M. Puaschunder, “Ein verhaltens-ökonomischer Ansatz zur Digitalisierung–Der Case der prinzipien-basierten Taxonomie.”