Acquisition, Modeling, and Evaluating Method of Driving Behavior Based on OBD-II: A Literature Survey

Galih Hermawan¹, Emir Husni²
¹Department of Informatics Engineering, Universitas Komputer Indonesia, Bandung, Indonesia
²School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, Indonesia.

Email: Galih.hermawan@students.itb.ac.id

Abstract. In this paper, we present the results of a literature survey related to driving behavior that aims to provide an overview of recent research related to driving behavior, including how to obtain data from onboard diagnostics (OBD-II), analyze data, model data, and evaluate models or systems. Driving behavior is closely related to driving habits that can be used for driver identification, driving characterization, improve fuel efficiency, and any other things. The contents of the discussion are arranged based on the research objectives, then will discuss the methods used and the evaluation results. Through this research is expected to provide insight into the acquisition and utilization of driving behavior for various things. Furthermore, provide an overview of research opportunities in the future.

1. Introduction

The World Health Organization (WHO) in 2018 reported that deaths from traffic accidents had increased by 1.35 million people per year. Nearly 3700 people died on the road every day all over the world. Some of the causes of accidents are poor vehicle safety standards, lack of law enforcement, people who are distracted or fatigued of driving, some are under the influence of drugs or alcohol, speeding, and failing to wear seat belts or helmets when an accident occurs [1]. To minimize accidents on the highway, WHO has created a technical manual called "SAVE Lives," which can be applied globally. The technical guidelines include speed management, leadership in road safety, infrastructure design and improvement, vehicle safety standards, traffic law enforcement, and survival after an accident [2].

One of the developed technologies in vehicles today to support the need for safety and comfort in driving is the Advanced Driver Assistance Systems (ADAS). It is a vehicle control system that uses environmental sensors (such as radar, laser, vision) to improve driving comfort and traffic safety by helping drivers recognize and react to potentially dangerous traffic situations [3]. In the last few decades, there have been several research related to strategy in modeling and improve driving monitoring and assistance systems. Research related to ADAS in general includes: development of fundamental driver assistance (such as: driving behavior models and highway safety), virtual testing and development environment (simulator), testing methods, sensors (radar, LIDAR, vision, and ultrasonic), data fusion and environmental representations (digital maps, maps location, and backend system), actuators (braking systems and steering wheel), engine and human interface (input device, visualization data, driver condition detection, and warning system), improvements stabilization (braking and steering stabilization), improved navigation capabilities on the road (parking assistance and lane change), and the development of other assistance systems in the future [4].

As previously mentioned, one of the fundamental studies on ADAS is related to driving behavior models. This study needs to be conducted because, based on WHO data, one of the essential factors involved in the accident is the driver’s factor [5]. Examples of driver factors are driving at excessive
speed, drunk, fatigue, driving in dark road conditions, poor vision, and vehicle condition factors. Furthermore, road conditions that make it difficult for drivers to see the road environment. There are four main topics related to driving, namely, the process of driving, driving disorders, fatigue, and how to drive [6]. We can use sensors to obtain information related to the condition of the driver, vehicle, and the environment (highway). The use of sensors includes sensors in the vehicle (such as brake, acceleration), additional sensors outside the vehicle (such as radar, LIDAR, laser), and physiological or biophysical sensors (such as eyes, head, and electroencephalography). In this paper, we conducted a literature survey related to driving behavior research, which focused on the needs of user identification, driving characterization, improved fuel efficiency, and several other applied examples.

This paper is organized as follows: Section 2 presents the research methodology and some basic definitions. Section 3 describes our literature survey on driving behavior research. Conclusions are drawn in Section 4, and future research opportunities are suggested.

2. Method
We used the literature survey research method in compiling this paper. The use of literature survey methods is essential to gain an in-depth understanding of the field of driving behavior research that already exists today. Research related to driving behavior only focused on data taken through OBD-II port devices. The data can be used directly or combined with other data, such as Global Positioning System (GPS) navigation, accelerometers, driver images, and environmental imagery.

On-Board Diagnostics (OBD) is an additional device in a vehicle that can monitor vehicle operational information on specific engine components, emissions controls, and components related to emissions. OBD-II is a second-generation OBD that was developed in 1992 [7]. Components or systems that are monitored by OBD include evaporation and emission control systems, actuators also sensors related to emissions, oxygen sensors, and any others. The initial OBD-II model was tested in vehicles in the United States in 1994, then had to be installed on all cars and small trucks in 1996 [8].

3. Results and Discussion
Driving behavior is related to the driving habits of a human when operating his vehicle. A person's driving style is generally carried out consciously. It is very likely also influenced by the current situation, such as weather conditions, road traffic, physical and psychological conditions of the driver, vehicle conditions, and various other factors. A person's driving behavior is different from other drivers [9].

Based on driving behavior, we can identify drivers, classify driving characteristics, predict travel time, identify types of roads, classify economic driving behavior models, and several other applied cases. Some studies have used data obtained directly from OBD-II, but some have used existing datasets, such as security datasets (ocslab) [10], hcilab [11], uah [12], and osf [13]. The next discussion is the results of our study regarding driving behavior.

3.1. Driver Identification
The motivation and benefits behind the driver identification research are to detect unauthorized access, determine driving service preferences, and help confirm the identity of the driver in the event of an accident or violation of the law [14].

In a study conducted by Bernardi et al. [15], in addition to analyzing driving behavior to identify drivers, they also analyzed to detect the type of road (lane) being passed. The paper proposed an approach to identify the driver and driver familiarity with the vehicle based on the driving behavior. The assumption is that a set of appropriate behavioral features can be used to recognize different drivers by capturing their different driving styles as well as to detect their familiarity with the car. The test result for driver identification has an accuracy value of at least 92%.

Research conducted by J. Zhang et al. [16] is by analyzing driving behavior based on the Ocslab dataset [10]. The method used for behavioral identification is a deep learning framework combining deep convolutional and Recurrent Neural Networks (RNN) in the paper is named DeepConvGRU and DeepConvLSTM. They used a framework that is based on a combination of repeated networks Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU)/Long Short-Term Memory (LSTM) to identify driving behavior using data CAN-BUS sensors in the vehicle. Within that
framework, GRU/LSTM cells are integrated into CNN to distinguish activities from the same driving behavior. DeepConvGRU/DeepConvLSTM based attention takes advantage of learning temporal dynamics. The experimental results show that the proposed framework can learn the features of the original signal and combining learning features without preprocessing certain. The F1-score of the test is equal to respectively 0.984 and 0.970.

Mekki et al. [14] analyzing data using four datasets are UAH-DriveSet [12], security driving dataset [10], osf [13], and hcilab [11]. The methods used are CNN and RNN/LSTM. Driver identification test results for the UAH-DriveSet dataset, Security-DriveSet, Multimodal OSF, and Hcilab respectively achieved an accuracy value of 89.86%, 95%, 62.12%, and 93.92%. In addition to testing the level of accuracy, it also addresses applied matters, namely by creating an interface that can display the driver identification results. Also included is the design of data processing and communication by utilizing edge computing, which can minimize latency when the driver identification stage is carried out in real-time.

Xun et al. [17] conducted experiments by reading and analyzing data from the OBD-II scan tool in real-time. The data obtained is processed using feature scaling and Principal Component Analysis (PCA). The algorithm used to identify the driver uses k-Nearest Neighbor (k-NN) and Naive Bayes combined with the voting mechanism. The test results involving data of ten drivers can achieve an accuracy value of 100%.

3.2. Driving Characteristic Classification
The physical condition of the driver while driving is very instrumental in determining safety and comfort within drive up. Several studies have been conducted to identify the driver's condition, whether in a state of drowsiness, fatigue, intoxication, or being out of focus. Driver condition identification can be done by observing head movements and facial features. If the driver's condition is deemed abnormal, as previously stated, the system can provide a warning message to the driver. On the other hand, the driver's abnormal behavior also allows being observed based on the driving behavior that is happening. Detecting driving behavior, in this case, is significant to improve driving safety.

Na Lin et al. [18] have collected various kinds of literature that discusses the classification of driving behavior, along with a general diagram of the system, the method used, and the results of each test. The following are examples of methods for classification, such as Neural Networks (NN), Hidden Markov Models (HMM), Fuzzy Control Theory, and Gaussian Mixture Model (GMM).

Chen et al. [19] model driving behavior to determine driver characteristics, which utilizes OBD-II data using the AdaBoost algorithm. This type of classification consists of two categories, namely driving safely or not safe. Sensor data used include vehicle speed, the engine rotation speed (RPM), gas pedal position (throttle), and calculated engine load. The results of classification testing using the AdaBoost algorithm can reach an accuracy value of 99.8%.

Zardosht et al. [20] conduct research related to behavior classification drive specifically only when the driver will make a turn. Data is taken directly through the OBD-II scan tool, including speed, vehicle features, gas pedal pressure, brake pedal pressure, steering wheel angle, and acceleration. The data is taken and analyzed when 5, 10, and 15 seconds before the bend. Twelve driver participants involved, where each drive for one hour. The obtained data is then analyzed and processed using the Hierarchical Clustering Analysis (HCA) algorithm. The test results show that there are two clusters with different characteristics. For cluster 1 (moderate driver): moderate speed, slightly increase the gas pressure when going to turn, brake gently, and gradually reduce speed; while the cluster 2 (aggressive driver): speed increases when going to maneuver turns, and speed reduction is performed quickly (suddenly).

3.3. Eco-Driving
With the increasing number of vehicles worldwide, will automatically increase the need for fuel consumption. The increasing use of fuel in motor vehicles will have an impact on environmental pollution (especially air) due to the removal of exhaust fumes from vehicles. So that motivation arises to save energy so that it can minimize air and environmental pollution, one of which is eco-driving. Eco-driving includes several driving methods that can achieve higher energy efficiency, thereby reducing fuel consumption and exhaust emissions.
Chen et al. [21], in his research, had utilized taxi operation data by considering the parameters of speed, acceleration, and duration. From the data obtained, then performed an analysis of fuel consumption. They used Principal Component Analysis (PCA) algorithm and multiple linear regression for feature analysis and modeling eco-driving driving behavior. The accuracy value of the testing model is 96.72%.

Ping et al. [22] have conducted studies related to the relationship between driving behavior and fuel consumption. Some methods used include clustering methods, object detection, and machine learning for data classification. In this study, they used a parallel spectral clustering algorithm to classify a dataset of driving signals collected from several drivers. The machine learning method used to model driving behavior is Long Short-Term Memory (LSTM).

3.4. **Applied Systems**
Some applied systems that involve driving behavior data are as follows.

3.4.1. **Online monitoring application.** A cloud website based application that can monitor the condition of vehicles and online driving behavior [23].

3.4.2. **VISCar.** An application that can monitor driving behavior and the state of the fuel, and can provide notification if the state of the fuel is in a small state while simultaneously showing the route to the nearest gas station [24].

3.4.3. **Driver Supervisor.** The application is designed for rental or travel cars that can monitor driving behavior in real-time. If the driving behavior is detected abnormally (or is considered to make a mistake), then through the Telegram bot can send notifications to the driver and owner of the vehicle at the same time [25].

3.4.4. **Anomaly Monitor.** A system that can detect if driving behavior is considered anomalous or abnormal. Then the system sends notifications to mobile devices where the data can be seen in real-time [26].

3.4.5. **FoRent.** A system that allows the owner and vehicle tenants can know the physical condition vehicles before and after being rented. In addition, in the mobile application, there are also features for registration, login, and vehicle position detection [27].

3.5. **Summary**
Regarding research methods, in general, each publication has set specific objectives. Driving behavior that is used for identification, regression, and classification (for which there is already a target value) uses the supervised learning algorithm. As for the clustering or data grouping without initial and target labels, then using the unsupervised learning algorithm.

OBD-II data can be used in full or even combined with other data if needed. Such as GPS, accelerometer, gyroscope, head pose the driver and others. Initial data, which is still in the form of raw data, usually must go through the stages of cleaning and normalization first. Every research has an output target that must be evaluated, which is generally done by measuring the accuracy value by comparing factual data and predictive data. Other assessment parameters that can be used are precision, recall, F1 score, MAPE, and others. In general, each study with different research objectives, has similar stages of data acquisition, data processing, data modeling, and evaluation, as shown in Figure 1.
From all stages of the study, as shown in Figure 1, using the same data, generally different at the stage of data processing and modeling. In data processing, the researchers used different approaches, such as feature extraction, data normalization, and the like. As for data modeling, it depends on the original purpose of the study, namely whether for data classification or clustering. This stage also includes the selection of machine learning algorithms in modeling data.

4. Conclusion
In the literature survey that we have conducted, there are many studies related to driving behavior. Existing research topics include modeling driving behavior, driver identification (fingerprint), classification of driver characteristics, determining the type of road, prediction of travel time, analysis of links to fuel consumption, to applied systems that can be used for daily use. For topics other than applied, research opportunities are still wide open in terms of increasing time efficiency and increasing accuracy (for example, in the case of identification, classification, test models). The researchers, until now, is still conducting more in-depth studies of the relationship between driving behavior and the driver's profile or preferences, environmental conditions, and certain types of vehicles.

Acknowledgement
We would like to thank the Indonesian Ministry of Education and Culture (Kemdikbud) through the BPPDN scholarship for funding our research. We also would like to thank Universitas Komputer Indonesia, who have helped and supported my study and this research.

References
[1] World Health Organization 2018 Global status report on road safety 2018 (World Health Organization).
[2] World Health Organization 2017 Save lives- a road safety technical package (World Health Organization).
[3] Gietelink, O., Ploeg, J., De Schutter, B., & Verhaegen, M. 2006. Development of advanced driver assistance systems with vehicle hardware-in-the-loop simulations. *Vehicle System Dynamics, 44*(7), pp. 569-590.

[4] Winner, H., Hakuli, S., Lotz, F., & Singer, C. (Eds.). 2016. *Handbook of driver assistance systems: Basic information, components and systems for active safety and comfort*. Springer International Publishing.

[5] Peden, M. 2004. World report on road traffic injury prevention.

[6] Khan, M. Q., & Lee, S. 2019. A comprehensive survey of driving monitoring and assistance systems. *Sensors, 19*(11), pp. 2574.

[7] United States Environmental Protection Agency. 2003. On-Board Diagnostic (OBD) Regulations and Requirements: Questions and Answer.

[8] Dzhelekariski, P., & Alexiev, D. 2005. Initializing communication to vehicle OBDII system. *ELECTRONICS, 5*, pp. 21.

[9] Miyajima, C., Nishiwaki, Y., Ozawa, K., Wakita, T., Itou, K., Takeda, K., & Itakura, F. 2007. Driver modeling based on driving behavior and its evaluation in driver identification. *Proceedings of the IEEE, 95*(2), pp. 427-437.

[10] Anon Driving Dataset - Hacking and Countermeasure Research Lab.

[11] Schneegass, S., Pfleging, B., Broy, N., Heinrich, F., & Schmidt, A. 2013, October. A data set of real world driving to assess driver workload. In *Proceedings of the 5th international conference on automotive user interfaces and interactive vehicular applications*, pp. 150-157.

[12] Romera, E., Bergasa, L. M., & Arroyo, R. 2016, November. Need data for driver behaviour analysis? Presenting the public UAH-DriveSet. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 387-392.

[13] El Mekki, A., Bouhoute, A., & Berrada, I. 2019. Improving Driver Identification for the Next-Generation of In-Vehicle Software Systems. *IEEE Transactions on Vehicular Technology, 68*(8), pp. 7406-7415.

[14] Bernardi, M. L., Cimitile, M., Martinelli, F., & Mercaldo, F. 2018. Driver and path detection through time-series classification. *Journal of Advanced Transportation, 2018*.

[15] Zhang, J., Wu, Z., Li, F., Xie, C., Ren, T., Chen, J., & Liu, L. 2019. A deep learning framework for driving behavior identification on in-vehicle CAN-BUS sensor data. *Sensors, 19*(6), pp. 1356.

[16] Xun, Y., Sun, Y., & Liu, J. 2019, May. An Experimental Study Towards Driver Identification for Intelligent and Connected Vehicles. In *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pp. 1-6.

[17] Lin, N., Zong, C., Tomizuka, M., Song, P., Zhang, Z., & Li, G. 2014. An overview on study of identification of driver behavior characteristics for automotive control. *Mathematical Problems in Engineering, 2014*.

[18] Chen, S. H., Pan, J. S., & Lu, K. 2015, March. Driving behavior analysis based on vehicle OBD information and adaboost algorithms. In *Proceedings of the international multiconference of engineers and computer scientists, 1*, pp. 18-20.

[19] Zardosht, M., Beauchemin, S. S., & Bauer, M. A. 2018. Identifying driver behavior in preturning maneuvers using in-vehicle CANbus signals. *Journal of Advanced Transportation, 2018*.

[20] Chen, C., Zhao, X., Yao, Y., Zhang, Y., Rong, J., & Liu, X. 2018. Driver’s Eco-Driving Behavior Evaluation Modeling Based on Driving Events. *Journal of Advanced Transportation, 2018*.

[21] Ping, P., Qin, W., Xu, Y., Miyajima, C., & Takeda, K. 2019. Impact of driver behavior on fuel consumption: classification, evaluation and prediction using machine learning. *IEEE Access*, 7, pp. 78515-78532.

[22] Husni, E., Hertanto, G. B., Wicaksono, D. W., Hasibuan, F. C., Rahayu, A. U., & Triawan, M. A. 2016, July. Applied Internet of Things (IoT): car monitoring system using IBM BlueMix. In *2016 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, pp. 417-422.

[23] Husni, E. 2017. Driving and Fuel Consumption Monitoring with Internet of Things. *International Journal of Interactive Mobile Technologies (iJIM), 11*(3), pp.78-97.
[25] Husni, E., & Hasibuan, F. 2018, September. Driver Supervisor System with Telegram Bot Platform. In *International Conference on Computational Collective Intelligence*, pp. 436-444.

[26] Amarasinghe, M., Kottegoda, S., Arachchi, A. L., Muramudalige, S., Bandara, H. D., & Azeez, A. 2015, August. Cloud-based driver monitoring and vehicle diagnostic with OBD2 telematics. In *2015 Fifteenth International Conference on Advances in ICT for Emerging Regions (ICTer)*, pp. 243-249.

[27] Saufi, N. N. C., Razak, N. S. M. A., & Mansor, H. 2019, January. FoRent: vehicle forensics for car rental system. In *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, pp. 153-157.