IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 26, 2019, accepted April 9, 2019, date of publication April 17, 2019, date of current version April 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2911704

Automotive ECU Data-Based Driver's Propensity Learning Using Evolutionary Random Forest

JONG-HYUN LEE[1], SANGMIN LIM[2], AND CHANG WOOK AHN[2], (Member, IEEE)
[1]Research Center for Convergence, Sungkyunkwan University, Suwon 2066, South Korea
[2]School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea

Corresponding author: Chang Wook Ahn (cwan@gist.ac.kr)

This work was supported in part by the GIST Research Institute (GRI) through the GIST in 2019, and in part by the Basic Science Research Program through the NRF funded by the Ministry of Education under Grant NRF-2017R1D1A1B03032785.

ABSTRACT Driving assistance systems in the automotive industry are constantly evolving and are already commercialized in various areas to provide consumers with safety and convenience. The recognition of driver's propensity is a key factor that can greatly affect the performance of such a driving assist system, but it still has numbers of technical limitations. This paper presents an evolutionary machine learning algorithm for recognizing driver's propensity by effectively learning a vast amount of ECU sensor data in the vehicle, and its performance is verified through system construction, data collection, analysis, and comparison test. The experiments showed that the proposed algorithm achieves a classification accuracy of 92.48% in a large amount of ECU data and reaches 7.03% higher accuracy than the average classification accuracy of existing classifiers. In addition, a scenario for a new safe driving assistance system is presented. The system can recognize the driver's propensity in real time using only the ECU information without attaching additional sensors, such as cameras and biometric information. It is expected that this system will help to recognize the driver's tendency shift, thereby inducing safe driving.

INDEX TERMS Big data learning, evolutionary computation, machine learning, evolutionary random forest, driving propensity recognition, vehicle safety assistant systems.

I. INTRODUCTION

The key to Big Data Learning is to quickly generate models while minimizing the loss of information behind the given data [1]. Previous studies on Big Data Learning often suggest advanced methods such as deep learning, yet those approaches have some constraints with optimization and calculation especially when applied to the actual industry [2]. However, the Random Forest algorithm [3], one of the ensemble algorithms based on a number of decision trees (weak classifiers), is effective when performing those tasks. Weak classifiers have a monotonous form which enables faster learning. Within this algorithm, weak classifiers also learn independently. Their structure is also convenient for thread-based distributed processing. Additionally, ensemble techniques such as boosting help to enhance learning accuracy and model expressiveness. In this paper, we propose an evolutionary approach to the Random Forest in order to learn and classify big data with a high temporal resolution from sensors. This study also aims to apply the devised algorithm to the mechanical data of vehicles.

Along with continuous technological developments in the present automotive industry, safety assistant systems have become an essential component of vehicles, and those systems have been widely deployed on a commercial scale. Other related safety support technologies such as pre-crash safety [4], forward vehicle collision warning [5], brake assist [6], blind spot warning [7], and lane keeping assist [8], previously mounted only in some selected models, are also prevalent in the global market. Nevertheless, those safety support systems and their related technologies are still limited in performance and efficiency.

Drivers' inclinations significantly affect the performance of safety support systems, but it is difficult to define and quantify the drivers' tendencies. There exist a countless number of driving patterns to be recognized; technical challenges arise, for example, when developing the vehicle safety support systems that can detect aggressive behavior. Furthermore, a given driver's tendency changes at any time. Although previous researchers attempted to measure driving

The associate editor coordinating the review of this manuscript and approving it for publication was Danilo Pelusi.

tendencies and define the drivers' inclinations by the Experience Index (EI) rating [9], sensor and vehicle driving data were rarely combined and examined. Accurate data from costly equipment like cameras and bio-signal sensors were missing in those previous studies.

In this paper, we build a driving pattern recognition algorithm for intelligent driver assistance systems. Mechanical signal data from electronic control units (ECU) in the vehicle such as vehicle acceleration, engine speed, torque, and steering wheel angle are collected by Control Area Network (CAN), and the combinations of those signal data are considered to reflect the driver's inclination. A technological convergence of evolution and machine learning techniques in this research provides a novel method which contributes to solve the challenges in the field of driving tendency recognition. Moreover, the evolutionary computation-based Random Forest helps to obtain higher accuracy in driving pattern classification. To validate our approach, we conduct comparison experiments between the proposed and traditional algorithms using real driving data collected by thirty expert drivers.

II. PROPOSED APPROACH TO EVOLUTIONARY MACHINE LEARNING FOR MASSIVE DATASET

Tendency recognition is a crucial element of an intelligent safety driving assistance system [10], [11]. However, it is very difficult to develop a learning algorithm with both high accuracy and flexibility for the tendency identification due to its complex and dynamic nature. For example, vehicle data from ECU that include almost all information regarding the operation of the vehicle may contain irrelevant information as well as the necessary ones. Additionally, tens of thousands of data showing the driver's propensity may emerge only a moment and suddenly vanish. Defining the optimal preprocessing function that preserves the critical information of the driver is also very challenging. Furthermore, if the size of raw data is too large, existing methods cannot effectively learn the pattern. Therefore, we devise a high-performance machine learning algorithm for a big and complex dataset by combining the evolutionary computation and conventional random forest.

A. EVOLUTIONARY RANDOM FOREST

The Random Forest Learning algorithm and its ensemble principle are widely used in handling big data, classifying, or processing pattern recognition. The algorithm constructs multiple weak-classifiers consisting of decision trees with randomly selected features or sample subsets of a training dataset. Each decision tree is learned by a given data subset and produces a decision-making model. In other words, features in their dataset are dismantled and examined to derive an adequate decision. In the classification process, the algorithm collects the predicted decision results from each model. Lastly, the majority voting rule makes the final prediction.

Despite the advantage of relieving the overfitting issue, the random rule of the Random Forest may impair the learning capability. Evolutionary computation which can improve the subset sampling process [11], [12], therefore, is expected to play a crucial role and complement the Random Forest by enhancing the search capability on the complex objective function.

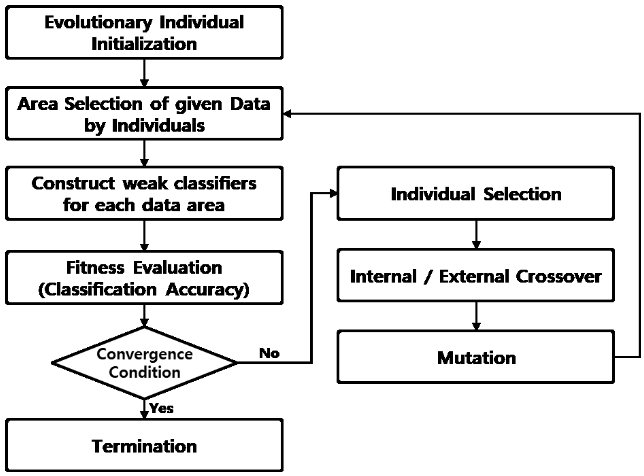

FIGURE 1. The overall process of evolutionary random forest algorithm.

1) THE OVERALL PROCESS OF EVOLUTIONARY RANDOM FOREST

As described in Figure 1, randomly generated rules at the initial stage select data subsets and assign those sets to each weak-classifier. The classifiers predict the class (i.e., driving tendency) of the training samples and take the majority votes to make a decision. The fitness of evolutionary individuals is evaluated by the results of classification accuracy. Then, the population of candidate individuals passes down the genetic traits and improves the accuracy by iterative operations of selection, crossover, mutation, and evaluation. If those individuals converge, the algorithm stops the repeating process and derives one optimal subset of individuals as a model for the classification. In the next step of using the trained model, it performs an ensemble classification with the Random Forest in regard to the optimal individual provided at the earlier stage.

2) THE STRUCTURE OF AN EVOLUTIONARY INDIVIDUAL

An individual is designed for the data subset selection in the tree structure. The tree has two types of nodes: *branch* and *leaf*. The leaf node gets the coordinates of the subset in the full dataset and denotes the column m and the row n as $\{n, m\}$. If necessary, an operator can choose all elements in a row or column. At first, the node gets assigned a random value within the size of data. After, it performs a search operation such as crossover or mutation. The branch node includes the union operator or the range operator. The comma symbol denotes the union operator which combines multiple references into one. The range operator, on the other hand, indicated by the colon symbol, produces one reference to every cell in between two references. The operators may constitute a region with the value of the child node. Sometimes, the subset region may cause incomplete problems for the features and samples. In such cases, the classifier uses the dataset from the complete area.

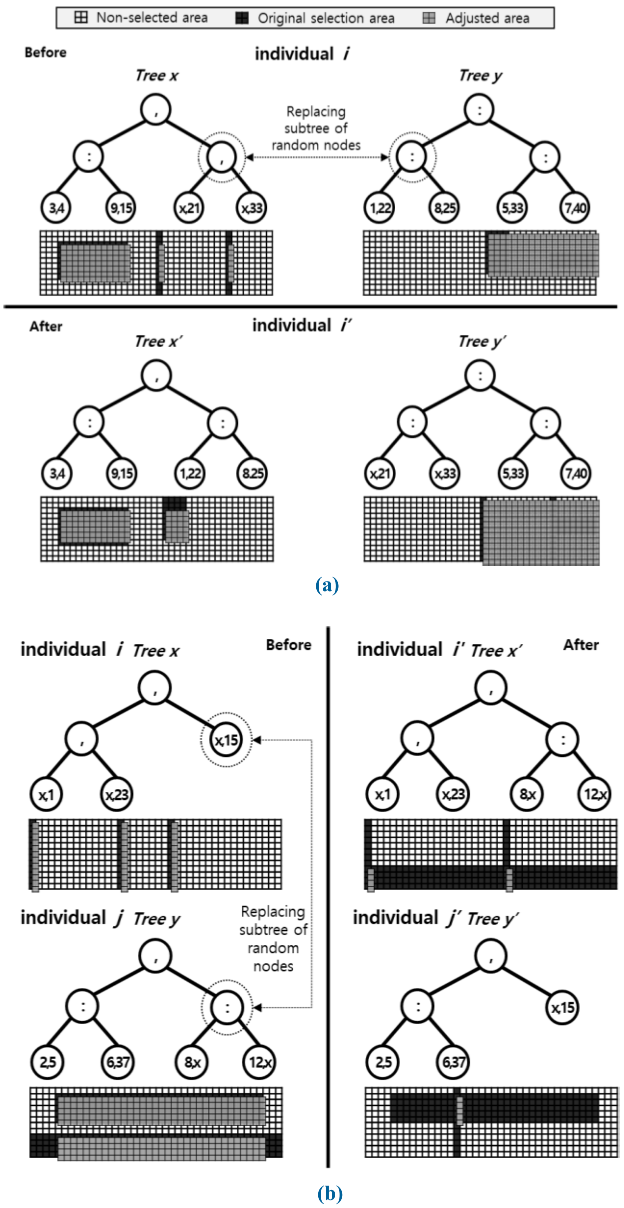

FIGURE 2. **Genetic operators for evolution: Internal and external crossover. (a) Internal crossover. (b) External crossover.**

3) GENETIC OPERATORS FOR EVOLUTION

Crossover, as in Figure 2, leads the interactions between the individuals or the trees in each individual. The number of tree-based individuals determines the type of crossover. During the external crossover, two subtrees from two different individuals are selected and swapped. The internal crossover conducts the swap in between the same individual. The selection and mutation processes follow the rule of the traditional tree-based evolutionary algorithms. Selection replicates better (i.e., fitter) individuals to the next generation through a competition. Mutation changes the value of any node from an individual to improve the search capability.

B. DRIVING TENDENCY RECOGNITION FOR ADVANCED DRIVER ASSISTANCE SYSTEMS

Unpredictability of the tendency appears throughout the drive. The signals are sometimes very clear, and sometimes they are too minuteness to recognize. For this reason, the tendency recognition system requires more accurate information from the sensor which directly connects with the driver.

We used a vehicle of Hyundai K9 in the driving experiment and examined the vehicle data gathered by CAN On-Board Diagnostic (OBD) device with a universal serial bus. Among a thousand ECU data, about one hundred sensor data including height, velocity, steering, gear, lane keeping, and fuel consumption were selected. The frequency of data collection is 1000 $\sim$ 2500 Hz. The entire raw data were utilized in the test to avoid the loss of necessary information. In order to collect various but consistent patterns, thirty driving experts traveled the same route of approximately 14km at a similar time of day as shown in Figure 3. Moreover, to obtain more accurate results, each driver conducted the tests twice so that anxiety associated with driving in an unfamiliar environment can be eliminated; only the second data were used in the experiment.

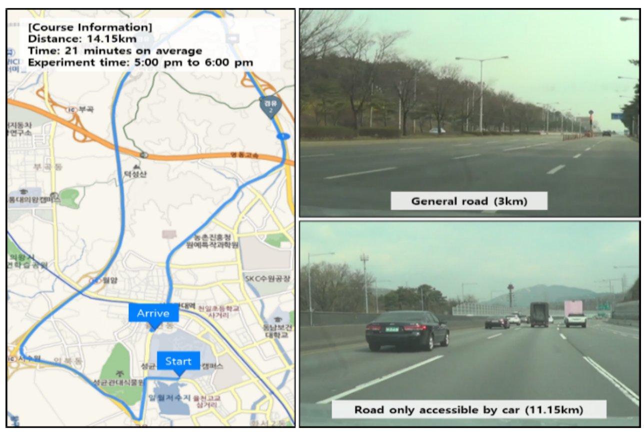

FIGURE 3. **Driving route information: Test examined under the same road conditions and at the same time of day.**

In the driving test (see Figure 4), we hired thirty male professional drivers (between the ages of 30 and 40) with at least five years of experience. All drivers were requested to drive as usual, and we did not inform them of any additional factors that can influence the test including the purpose of the experiment. As a result, datasets of hundreds of attributes and 75 million instances of gigabytes size were gathered. For the effective learning and classification of the collected data, we devised the new method that corresponds to the

FIGURE 4. 30 expert drivers participated and travelled the same route twice.

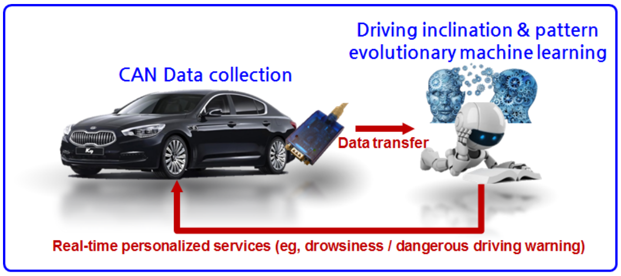

FIGURE 5. Vehicle data and driver pattern recognition system for real-time personalized services.

characteristics of these data and conducted comparison experiments with conventional methods.

C. VALIDATION WITH ACTUAL DRIVING DATA

Each driver's distinct driving tendency is evaluated by the statistical values from the actual driving data of lane involvement, over-speeding, or sudden starts and stops. Based on the safe driving instructions of the traffic agency, each one received a risk score; high-risk drivers were the ones who violated such instructions, and the others who kept those regulations got a better safety rating. In this way, a learning model for risk driving can be generated and used to rate other drivers, and Figure 5 depicts such a configured system. Since the accuracy of the learning model is the key factor to classify driving patterns, the experiments in the following section focus mainly on the learning accuracy.

III. EXPERIMENTAL RESULTS

The test environment of the proposed method was constructed in C++. We set the parameters of the number of individuals, depth limitations of the tree in individuals, iteration, and probabilities of crossover and mutation as 100, 4, 100 generations, 0.7 and 0.1, respectively. Additionally, the maximum depth of decision trees is unlimited, and the number of decision

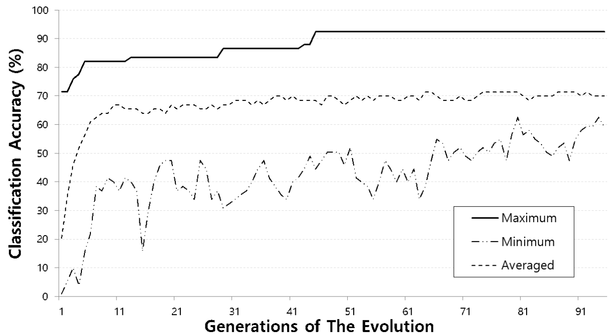

FIGURE 6. Maximum, minimum, and average classification accuracy (%) in terms of generations.

TABLE 1. Classification accuracy (%) of the conventional algorithms and the proposed ERF.

	Accuracy
SVM	88.57%
MLP	86.71%
Adaboost M1	83.72%
CART	82.82%
Random Forest	85.43%
Evolutionary Random Forest	92.48%

trees is 100. Figure 6 shows the accuracy variation by the proposed Random Forest in the learning process through the generations. Although the early generations offered low accuracy due to the incompleteness of the data sets, the evolution of tree subset selectors in later generations enhanced the average, maximum and minimum classification accuracy of one hundred individuals (i.e., population). Therefore, we can confirm that the proposed approach is helpful for refining the classification accuracy of the learning method.

Conventional classifiers such as Support Vector Machine (SVM) [13], Multilayer Perceptron (MLP) [14], AdaBoost. M1 [15], Classification and Regression Trees (CART) [16] of a decision tree, and the Random Forest were added in the classification experiment to analyze the accuracy through comparison. The parameters of those classifiers were set at a standard value of WEKA [17], and the accuracy results were obtained by 10-fold cross-validation.

Table 1 displays the classification accuracy of the conventional classifiers (implemented in WEKA) and that of the devised method. Our classification technique achieved 7.03% higher accuracy, compared to the average of other classifiers, and 3.91% higher than SVM, which has the best performance among the traditional methods.

Additional comparative experiments were conducted to verify the performance of the ERF in terms of feature selection. We adopted a subset based feature selection algorithm to search for the smallest subset with little or no loss in accuracy, and then compared the values obtained by the subset to the values from the whole set. The algorithm for evaluating the subset uses the Random Forest, and the parameters of tree depth and the number of trees are identical.

Initially, we selected 63 features from hundreds of CAN data to make the experiment simple. However, even with the reduced 63 features, the total number of possible combinations, 2 to the 63rd power (=9.22337e + 18), is almost uncountable. In this sense, we used the Best First Search (BFS) in order to search for a combination of the subsets. The BFS is a sort of greedy algorithm based on backtracking and hill climbing. It has been widely used in many applications due to its fast speed and simple structure.

Table 2 summarizes the results of the fifteen best features extracted by the BFS. Note that ERF still gave the highest accuracy of 91.74% with the reduced data. Among other existing algorithms, the Random Forest recorded 88.47% accuracy which displays the highest increase of 3.04%.

TABLE 2. Classification accuracy (%) of the fifteen best features by the BFS and difference values w.r.t. Table 1.

	Accuracy	Difference w.r.t. Table 1
SVM	87.42%	-1.15%
MLP	83.70%	-3.01%
Adaboost M1	83.96%	0.24%
CART	83.28%	0.46%
Random Forest	88.47%	3.04%
Evolutionary Random Forest	91.74%	-0.74%

TABLE 3. List of the fifteen features from the BFS.

Selected Features	
1	Height of front right corner
2	Height of rear left corner
3	Height of rear right corner
4	Vehicle acceleration
5	Mileage counter for odometer, Left hand
6	Engine speed
7	Standard torque ratio
8	Fuel consumption
9	Steering wheel angle
10	Current Gear
11	Vehicle speed calculated by TCU
12	RR Wheel Pulse Count
13	Left Lane Departure Warning Status
14	Right Lane Departure Warning Status
15	Status of Recognition of Left Line

Table 3 shows the results of extracting the minimum features with little or no loss of accuracy through the BFS method. The features for the driver classification listed in Table 3 as the objective function revealed the characteristics of what is known to be the actual driving propensity.

To test whether the selected features comprise the optimum or not, we performed additional experiments by removing features from the given list. The experiments were based on the Random Forest because of its highest increase rate of accuracy. Table 4 shows the accuracy measurement results after deleting the corresponding individual attribute from the fifteen attributes in Table 3.

TABLE 4. Random forest's classification accuracy (%) after removing each attribute and difference values w.r.t. Table 2.

Removed individual attribute	Random Forest Accuracy	Difference w.r.t. Table2
Height of front right corner	82.23%	-6.23%
Height of rear left corner	81.33%	-7.13%
Height of rear right corner	79.75%	-8.72%
Vehicle acceleration	87.12%	-1.35%
Mileage counter for odometer, Left hand	88.24%	-0.22%
Engine speed	87.09%	-1.37%
Standard torque ratio	81.21%	-7.26%
Fuel consumption	87.49%	-0.97%
Steering wheel angle	87.68%	-0.79%
Current Gear	87.49%	-0.97%
Vehicle speed calculated by TCU	87.11%	-1.36%
RR Wheel Pulse Count	88.31%	-0.16%
Left Lane Departure Warning Status	87.46%	-1.01%
Right Lane Departure Warning Status	87.79%	-0.67%
Status of Recognition of Left Line	87.76%	-0.71%

TABLE 5. Random forest's classification accuracy (%) after removing a group of attributes and the difference values w.r.t. Table 2.

Removed partial attribute combinations		Random Forest Accuracy	Difference w.r.t. Table2
Group 1	Fuel consumption	87.39%	-1.07%
	Left Lane Departure Warning Status		
	Status of Recognition of Left Line		
Group 2	Fuel consumption	86.77%	-1.70%
	Vehicle speed calculated by TCU		
	Left Lane Departure Warning Status		
	Status of Recognition of Left Line		
Group 3	Left Lane Departure Warning Status	87.98%	-0.49%
	Right Lane Departure Warning Status		
Group 4	Mileage counter for odometer, Left hand	87.78%	-0.69%
	Fuel consumption		
	Left Lane Departure Warning Status		
	Right Lane Departure Warning Status		

As described in Table 4, the accuracy decreased if any of the attributes from the original list were removed. After simulating all possible cases, we conclude that any subset of features in Table 3 cannot provide more accurate results. In other words, the features listed in Table 3 cannot be further optimized by eliminating any of them; they already formed an optimal solution.

Although it was not possible to enhance the accuracy further, the combination of two or more attributes deletion, in some cases, decreased the accuracy no greater than 2% (see Table 5). For example, the accuracy with 11 attributes in Group 4 offered a similar level of the values with the fifteen features and better than the results from all 63 combinations. Such an outcome indicates that the change in classification accuracy due to the removal of features is made up of a

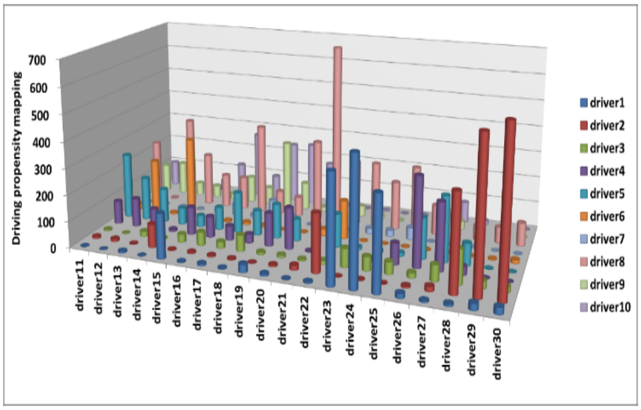

FIGURE 7. **Mapping new 20 drivers' (Driver 11 ∼ 30) data to the driving propensity models from 10 drivers (Driver 1 ∼ 10).**

combination rather than a simple sum. Obviously, discovering the optimal feature combination is very hard. Nevertheless, the task is worth considering further in the future since a very high accuracy can be achieved with the minimal data.

Note that our attempts to raise the accuracy (of the Random Forest) through feature removal and to perform the feature optimization in this experiment actually gave us a lower accuracy than that from the ERF. This implies that conventional approaches have a limit in performance, but the sample selection performed by the suggested ERF is very effective.

IV. DANGEROUS DRIVING DETECTION SYSTEM USING EVOLUTIONARY RANDOM FOREST

This section introduces a service application of the proposed approach. The target system is a vehicle operation risk analysis system using ECU without any additional device. The system consists of a combination of the driver's inclination, Experience Index (EI) rating, and the proposed Evolutionary Random Forest. The overall process is as follows: The drivers' data to be learned are modeled by the ERF, and the EI rating is used to compute the risk scores of the drivers. Then, a new driver's data is analyzed based on the driver models (as a reference) and the driver gets its own risk score.

Figure 7 shows the results of applying the driving propensity model obtained on drivers 1-10 to un-trained data on drivers 11-30. Each new driver could be mapped to two to four models in the above graph, and the mapping ratio of the top three models was from 60% to 95%. These results suggest that each driver has a unique operating pattern; however, similar patterns exist in part, and recognizing such patterns is viable in such an environment.

From the ECU data, we analyzed the driving propensity for the ten modeled drivers, as in Table 6. The risk score is a risk indicator of the EI rating that is calculated with four factors: the average vehicle speed, the number of times of steering (more than 15 degrees), the number of lane changes, and the maximum speed. If a participant's record exceeds the median value in each factor, then the risk score is increased by one. Therefore, the risk score of 'driver i' is defined as $risk\_score(i) = \sum_{k=1}^{4} u(x_{i,k} - \bar{x_k})$, where $\bar{x_k}$ is the median value of the kth factor of drivers, $x_{i,k}$ is the data of the kth factor of driver i, and $u(x)$ is a unit step function. Considering the risk scores of the drivers 1-10, as shown in Table 6, four points are given to the driver 10, three points for the drivers 3 and 9, two points for the drivers 1, 5, 6, and 8, and one point for drivers 2, 4, and 7.

Figure 8 shows the results of the risk score estimation for un-trained drivers on the basis of the reference models learned from ten drivers; four un-trained drivers (drivers 11, 17, 21, and 28) have been chosen for a closer look. For instance, 'driver 13' has the propensity of aggressive and dangerous driving, 'driver 21' has the propensity of gentle and safe driving, and 'driver 17' and 'driver 28' tend to often drive a car in a dangerous way, according to the given situation.

TABLE 6. **EI rating analysis of the driving propensity model on ECU data of 10 drivers.**

	Driver1	Driver2	Driver3	Driver4	Driver5	Driver6	Driver7	Driver8	Driver9	Driver10
Std. Yaw rate (degree/sec.)	0.9463	0.9605	0.9982	1.1084	1.0017	0.8270	0.9985	1.0447	1.1543	1.1139
Diff. Min. Max. Yaw rate	9.95	9	10.98	11.58	10.4	7.6	11.03	10.49	11.96	12.44
Session vehicle speed	58.332	61.580	55.167	64.370	59.341	60.796	58.944	60.136	67.401	62.146
Avg. vehicle speed	62.133	61.580	65.557	64.370	62.127	62.266	62.884	63.162	71.047	66.665
Standard torque ratio	0.4436	0.3744	0.2631	0.4684	0.4962	0.1322	0.5818	0.4124	0.1974	0.4801
Avg. Fuel consumption (liter/sec)	0.0031	0.0030	0.0024	0.0030	0.0026	0.003	0.0030	0.0022	0.0028	0.0026
Total Fuel consumption (liter)	2.6915	2.4425	2.1568	2.2954	2.1738	2.2859	2.5165	1.8163	2.0319	2.0402
No. accelerator pedal	32	38	36	39	44	36	27	52	30	50
Filtered Accelerator Pedal Value	37.291	34.705	45.123	41.251	37.749	38.793	40.032	42.794	42.625	49.658
Std. Accelerator Pedal Value	20.795	22.162	27.707	21.885	22.295	24.801	21.390	29.788	28.506	34.271
No. steering wheel (above 15 deg.)	78	73	63	79	79	83	67	91	82	91
Avg. gears during driving	6.3497	6.8276	5.9269	6.9839	6.5846	6.6438	6.5170	6.4768	6.9098	6.2816
Number of lane changes	8	9	9	8	10	4	9	5	7	13
Driving time (sec.)	850	789	870	761	825	762	830	814	721	782
Maximum speed (km/h)	105	95	99.63	88.93	85.9	91.72	85.86	86	105.81	114.71
Risk Score	**2**	**1**	**3**	**1**	**2**	**2**	**1**	**2**	**3**	**4**

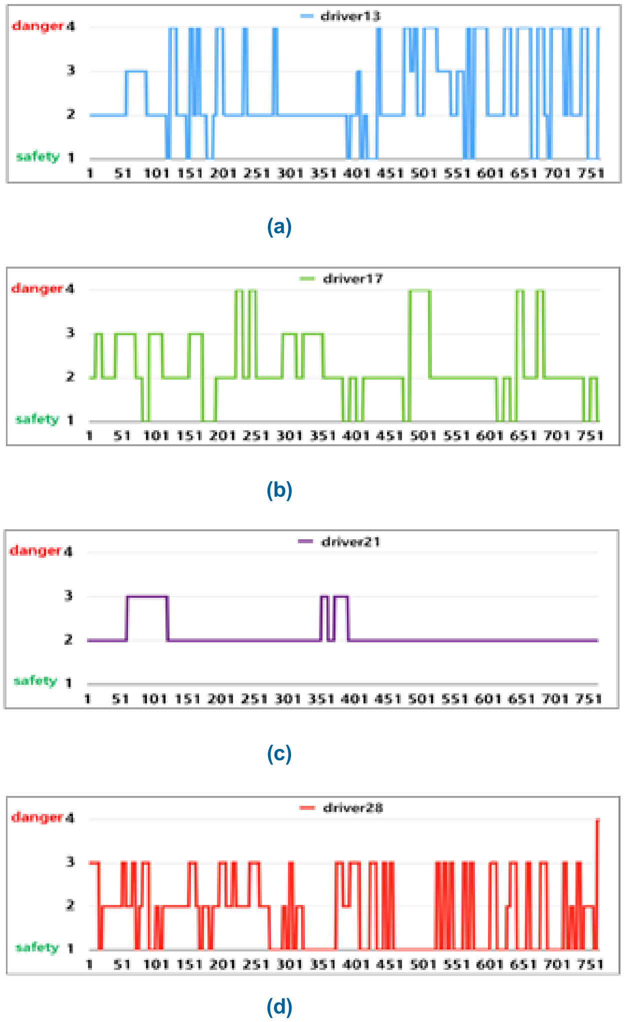

FIGURE 8. Results of risk score estimation for un-trained drivers. The horizontal axis is the running time (sec), and the vertical axis is the risk score. (a) Driver 13. (b) Driver 17. (c) Driver 21. (d) Driver 28.

Last but not least, the proposed approach can model a large amount of data and obtain driver's propensity using only the ECU machine data. In addition, the model can be utilized for providing proper feedback for the driver of risk driving patterns. However, further studies on the reference driver model are needed to reflect more diverse, real-world situations.

V. SUMMARY

In previous studies, recognizing driving tendencies has been a very challenging task. Thus, this paper focuses on the recognition accuracy of driver's propensity and proposes a new learning method based on the ERF. The suggested method promotes the search capability by the nature of evolution. With experiments, we validated the proposed technique and its superiority in comparison to the conventional classifiers in the selected dozens of sensor data from ECU.

Moreover, we constructed an environment to gather machine data by accessing the ECU of the vehicle. Then, we collected actual driving data of multi-specialty drivers. With the data, we performed a task of analysis and modeling, and then designed an actual application system. Since there are many barriers when applying the related theory to such real-world problems, earlier researches omitted such processes. However, we have successfully obtained meaningful empirical results. The proposed dangerous operation detection system is cost-effective because it does not require any additional equipment, such as cameras and sensors.

Future studies may require an improved structure of selecting area operators. Also, adaptive system design can be considered to mitigate the sensitivity of parameter settings. Additionally, performance benchmarking and optimization using additional data for the offered scheme is needed. It is expected to verify the operation model based on the risk scores in the application step. In the next step, we plan to apply our method to other applications with similar characteristics, such as in-vehicle services tailored to the driver's emotion.

REFERENCES

[1] P. Simon, *Too Big to Ignore: The Business Case for Big Data*. vol. 72. Hoboken, NJ, USA: Wiley, 2013.

[2] X.-W. Chen and X. Lin "Big data deep learning: Challenges and perspectives," *IEEE Access*, vol. 2, pp. 514–525, 2014.

[3] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[4] E. Infantes *et al.*, "The effect of pre-crash safety systems to occupant protection in offset frontal impacts," SAE Tech. Paper 2015-26-0164, 2015.

[5] K. D. Kusano and H. C. Gabler, "Safety benefits of forward collision warning, brake assist, and autonomous braking systems in rear-end collisions," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1546–1555, Dec. 2012.

[6] A. Badea-Romero *et al.*, "Assessing the benefit of the brake assist system for pedestrian injury mitigation through real-world accident investigations," *Saf. Sci.*, vol. 53, pp. 193–201, Mar. 2013.

[7] B.-F. Wu, H.-Y. Huang, C.-J. Chen, Y.-H. Chen, C.-W. Chang, and Y.-L. Chen, "The blind spot warning system for daytime and nighttime," in *Proc. IEEE Int. Conf. Ind. Technol. (ICIT)*, Mar. 2012, pp. 591–597.

[8] K. Mineta, K. Unoura, and T. Ikeda, "Development of a lane mark recognition system for a lane keeping assist system," SAE Tech. Paper 2003-01-0281, 2003.

[9] D. Yagil, "Interpersonal antecedents of drivers' aggression," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 4, no. 2, pp. 119–131, 2001.

[10] X. Wang and J. Zhang, "Extraction and recognition methods of vehicle driving tendency feature based on driver-vehicle-environment dynamic data under car following," *Int. J. Comput. Intell. Syst.*, vol. 4, no. 6, pp. 1269–1281, 2011.

[11] J.-H. Lee and C. W. Ahn, "An evolutionary approach to driving tendency recognition for advanced driver assistance systems," in *Proc. MATEC Web Conf.*, vol. 56, 2016, p. 4.

[12] D. B. Fogel, *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*, vol. 1. Hoboken, NJ, USA: Wiley, 2006.

[13] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.

[14] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences," *Atmos. Environ.*, vol. 32, nos. 14–15, pp. 2627–2636, 1998.

[15] E. A. Cortés, M. G. Martínez, and N. G. Rubio, "Multiclass corporate failure prediction by Adaboost. M1," *Int. Adv. Econ. Res.*, vol. 13, no. 3, pp. 301–312, 2007.

[16] W.-Y. Loh, "Classification and regression trees," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 1, no. 1, pp. 14–23, 2011.

[17] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.

JONG-HYUN LEE received the B.S., M.S., and Ph.D. degrees from Sungkyunkwan University, in 2010, 2012, and 2017, respectively, all in computer science. He is currently a Postdoctoral Researcher with the IT Convergence Research Institute, Sungkyunkwan University. Since 2017, he has also been the CEO of CreativeMind Inc. His research interests include artificial intelligence, machine learning, evolutionary algorithm, and swarm robotics.

SANGMIN LIM received the B.A. degree in economics from Northwestern University. He is currently pursuing the master's degree with the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST), South Korea. Prior to his graduate studies at GIST, he involved in IT consulting, accounting, investment banking, and software development. His primary research interests include artificial intelligence and machine learning in finance.

CHANG WOOK AHN received the Ph.D. degree from the Department of Information and Communications, GIST, in 2005. From 2005 to 2007, he was with the Samsung Advanced Institute of Technology, South Korea. From 2007 to 2008, he was a Research Professor at GIST. From 2008 to 2016, he was an Assistant/Associate Professor with the Department of Computer Engineering, Sungkyunkwan University (SKKU), South Korea. He is currently a Professor with the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST), South Korea. His research interests include genetic algorithms/programming, multi-objective optimization, evolutionary neural networks, and quantum machine learning.