Avrupa Bilim ve Teknoloji Dergisi
Sayı 20, S. 774-782, Aralık 2020

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
No. 20, pp. 774-782, December 2020

Research Article

Performance Evaluation of Major Classification Algorithms for Aggressive Driving Detection using CAN-bus Data

Berat Karabuluter[1], Özgür Karaduman[2*], Murat Karabatak[3], Haluk Eren[4]

[1] Firat University, Faculty of Engineering, Departmant of Software, Elazig, Turkey, (ORCID: 0000-0002-8712-3382), beratkarabuluter@gmail.com
[2*] Firat University, Faculty of Engineering, Departmant of Software, Elazig, Turkey, (ORCID: 0000-0002-6569-3616), okaraduman@firat.edu.tr
[3] Firat University, Faculty of Technology, Departmant of Software, Elazig, Turkey, (ORCID: 0000-0002-6719-7421), mkarabatak@firat.edu.tr
[4] Firat University, School of Aviation, Air Traffic Management, Elazig, Turkey, (ORCID: 0000-0002-4615-5783), heren@firat.edu.tr

(First received 29 May 2020 and in final form 6 December 2020)

(**DOI:** 10.31590/ejosat.743076)

ATIF/REFERENCE: Karabuluter, B., Karaduman, Ö., Karabatak, M. & Eren, H. (2020). Performance Evaluation of Major Classification Algorithms for Aggressive Driving Detection using CAN-bus Data. *European Journal of Science and Technology*, (20), 774-782.

Abstract

Detection of driver moods associated to driving style such as drowsy, distracted, vigilant, calm, or aggressive driving is one of the main problems of Advanced Driver Assistance Systems and it obviously plays vital role in the prevention of traffic accidents. The main goal of this study is to compare the performances of major Supervised Learning based Classification Algorithms (SLCAs) for aggressive driving detection, which is one of the fundamental problems for understanding driver mood or driving style through CAN (Control Area Network) bus sensor data. These algorithms utilize CAN-bus data acquired by OBDII (On-board Diagnostics) socket of the vehicle. In our experiments, to get ground truth data, many trials referring to aggressive and calm driving have been conducted by different subject drivers and these sensor data have been labeled as "aggressive" and "calm". Afterwards, these transformed into training data to assess performances of SLCAs. As a result, the Naïve Bayes Classifier has been found to be more successful than the others.

Keywords: CAN-bus, Driving safety, Aggressive driving, Intelligent vehicles, Classification algorithms.

CAN-bus Verileri Kullanarak Agresif Sürüş Tespiti için Temel Sınıflandırma Algoritmalarının Performans Değerlendirmesi

Öz

Uykulu, dikkati dağılmış, dikkatli, sakin ya da agresif sürüş gibi sürüş tarzıyla ilişkili olan sürücü ruh hallerinin tespiti İleri Sürücü Destek Sistemlerinin (İSDS) ana problemlerinden biridir ve trafik kazalarının önlenmesinde hayati rol oynamaktadır. Bu çalışmanın temel amacı, CAN-bus sensör verilerini kullanarak sürücü ruh halini veya sürüş tarzını anlamanın temel problemlerinden birisi olan agresif sürüş tespiti yapmak için Eğiticili Öğrenme Tabanlı Sınıflandırma Algoritmalarının (EÖSAs) performanslarını karşılaştırmaktır. Bu algoritmalar, aracın OBDII soketinden elde edilen CAN-bus verilerini kullanır. Deneylerde, referans verilerini elde etmek için agresif ve sakin sürüşe ilişkin bir çok deneme sürüşü farklı sürücüler tarafından gerçekleştirilmiştir. Elde edilen veriler "agresif" ve "sakin" olarak etiketlenmiştir. Ardından, EÖSA' ların performansını değerlendirmek üzere eğitim verilerine dönüştürülmüştür. Yapılan performans değerlendirmesi sonucunda, Naïve Bayes sınıflandırıcısının diğerlerinden daha başarılı olduğu görülmüştür.

Anahtar Kelimeler: CAN veriyolu, Sürüş güvenliği, Agresif sürüş, Akıllı araçlar, Sınıflandırma algoritmaları.

* Corresponding Author: Firat University, Faculty of Engineering, Department of Software Engineering, Elazig, Turkey, ORCID: 0000-0002-6569-3616, okaraduman@firat.edu.tr

1. Introduction

1.1. Motivation

Every year in the world, thousands of traffic accidents occur, resulting in many injuries and deaths. These accidents many of which result in loss of life and property are often caused by driver mistakes. According to NHTSA and Vantage Auto Club data, 66% of fatal traffic accidents are caused by "aggressive driving" (safemotorist, iii.org, aaafoundation). For example, according to the statistics of the Turkish Statistical Institute and Research Department, 313.746 out of the 1.229.364 traffic accidents, which result in death and injury in Turkey in 2018, namely 89.6%, are caused by the drivers (tuik.gov.tr). About 55% -70% of these mistakes are driver mistakes which can be defined as "aggressive driving". According to aforementioned information, it can be concluded that driving styles and driver moods play an important role in traffic accidents.

1.2. Problem Statement and Proposed Approach

In the most general sense, "aggressive driving" can be defined as performing dangerous driving with risk of accident by making sudden speed or position changes that violate traffic rules. It is generally assumed that aggressive driving is in the form of low gear driving, sudden braking, sudden acceleration, sudden departure, sudden and frequent gear shifting, and uncontrolled lane change. Therefore, the real problem here is that sudden driving movements of the driver can be detected, which can represent a "driving mood" in a means. The existing mood of driver maybe different from his/her driving style. This situation looks like a player in a theatre. His/her existing mood can be happy but he/she plays like in unhappy person role. In a similar manner happy driver can drive aggressive. In opposite, an unhappy driver can drive calm in response to his/her driving experience or professionalism. Without acquiring and processing driving data on steering wheel, driving style of a driver cannot be understood. This state can be considered as "driving mood" rather "driver mood". As shown in Figure 1, State C is the impulsive response of State A. Likewise, State E is the impulsive response of State C. In this study, to detect these movements it has been used the OBDII information of the vehicle received via CAN bus, which includes the vehicle's internal sensor information. By means of the built-in sensors, it can be obtained clues for detecting driving moods thanks to the driver's pedal movements and associated information such as speed, acceleration, and RPM (Revolution per Minute) of the vehicle.

In this study, external sensors such as cameras, radars, or LIDARs are not utilized. Therefore, the cost of problem is tried to be minimized solely using vehicle internal sensors. Specified classification algorithms are employed to detect aggressive driving via the CAN bus data. Data acquisition is carried out through CAN bus on a designated parkour.

There are three stages of moods referring State A, State C, and State E in Figure 1. Also, there are two stages affecting mentioned moods addressing State B and State D. We assume driver's native mood looks like a system. If an impulse like State B is applied to the system, State C is obtained. Then, State D is applied to State C. The final one is the responsive state (State E) to previous impulsive state (State D). Two different types of driving have been experimented, which are labelled as "aggressive driving" and "calm driving".

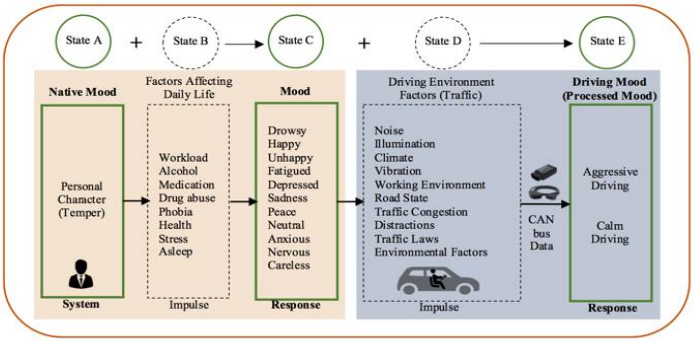


Figure 1. The story of driving behavior.

Performance of the deployed algorithms has an important role in estimating driving style. As a matter of this fact, it needs to be assessed specified algorithms. The proposed approach allows researchers to find best algorithm for related problem. In this study, some well-known methods have been deployed, which are ANN, Naïve Bayes, SVM, C4.5 (J48), and K-Nearest Neighbor algorithms.

1.3. Related Works and Contributions

In the literature, many sensors have been used for determining driving style, one of which is external camera. Ö. Kumtepe et al. introduced a camera based system to detect driving style (Kumtepe at al, 2015; Kumtepe at al, 2016). They obtained aggressive driving rate using aforementioned components as nearly 90%. There are many driving behaviour estimation studies based on smartphone sensors in literature (Trivedi at al, 2011; Eren at al., 2012; Bergasa at al., 2014; Koh at al., 2015; Li at al., 2016; Oliver at al., 2019). One of them is that of Bergasa et al. they were exploited by smartphone sensor data such as GPS sensor, acceleration sensor, and its camera to obtain aggressive driving level. Lane weaving and drifting rates are found, and aggressive driving scoring is performed. Authors in their study get a performance over 82%. Moreover, Fuzzy classification method is preferred in some of studies (Imkamon at al, 2008; Wu at al, 2012; Songkroh at al, 2014; Fazio at al, 2016; Arefnezhad at al, 2019; Wesseleny at al,2019). Determination of aggressive driving without CAN bus data is achieved by the studies in (Waitkus at al, 2014; Li at al, 2014; Vignali at al, 2019; de Naurois at al, 2019). Simulator based studies rather using real vehicles exist in pressed resources introducing criteria for aggressive driving (Doshi at al, 2010; Gregoriades at al, 2013; Shirazi at al, 2014; Keklikoglou at al, 2018). Some of studies comprise influence of aggressive driving estimation on emission (Dia at al, 2015; Sun at al, 2016; Stogios at al, 2019; Faria at al, 2019).

OBDII sensor data is also involved in some of the aforementioned studies. Besides, attribute selection for CAN bus data reduction is conducted (Karaduman at al, 2013; Taylor at al., 2015; Fugiglando at al., 2018; Lokman at al., 2019; Le at al., 2020). The proposed study introduces an approach comparing calm and aggressive driving classification algorithms and exploits internal sensors existing in subject vehicle. As known, most of the vehicles have different kind of internal sensors. However, we have preferred common sensors existing in most of vehicles. To cope with this main problem, we have tried to use minimum number of sensors. Therefore, this scheme brings cost efficiency, portability, and compactness due to not using external sensors such as camera, radar, LIDAR, GPS, and smartphone sensors. In this study, some training based methods are used by real experimental subjects on a specified route, which are ANN, Naïve Bayes, SVM, C4.5, and

KNN. This comparative approach tries to find best supervised algorithm for aggressive and calm driving detection problem.

1.4. Paper Organization

Section II provides comprehensive details about methodology including Data Acquisition, Normalization, Learning and Classification Algorithms, Validation and Performance Evaluation. Section III comprises Experimental Results, subsections of which consist of data collection via CAN bus, estimation results of normalization, performance measure of selected algorithms, validation and performance evaluations. Conclusion and future works are in Section IV.

2. Material and Method

While determining driving style of a motorist, a system diagram with all the stages including performance comparison of potential classification methods and utilization of internal sensors' information via CAN bus are provided in Figure 2.

In the first stages, training data are collected, which includes aggressive and calm driving. The second stage comprises normalization process. In the stage three, the data labelled as aggressive/calm are trained by specified algorithms, which are ANN, SVM, KNN, C4.5, and Naïve Bayes. The last stage includes performance comparisons for aforementioned algorithms supporting with validation.

2.1. Data Acquision

The main goal of the study is to assess performances of driving style detection for specified classification algorithms using OBDII sensor data. All the data are collected on a specified route. Two types of individual experiments representing aggressive and calm driving are conducted to get ground truth data. In this manner, two different data tables are obtained. For the classification, all the sensor data are not involved in the process. Hence, specified attributes are selected from OBDII sensor data as shown in Table I. The other sensor data are not selected, since they reveal same characteristics for both aggressive and calm driving, in other words, they don't expose discriminative features

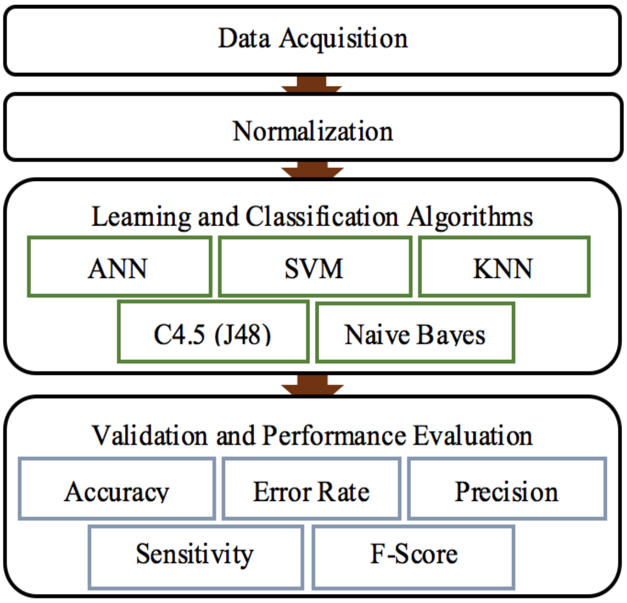


Figure 2. Stages of the proposed system

Table 1. The selected sensor data model

	Sensor Data Type					
Number	*A1*	*A2*	*A3*	*A4*	*A5*	...
X1	$A1_1$	$A2_1$	$A3_1$	$A4_1$	$A5_1$	
X2	$A1_2$	$A2_2$	$A3_2$	$A4_2$	$A5_2$	
X3	$A1_3$	$A2_3$	$A3_3$	$A4_3$	$A5_3$	
.	.	.	.	.	.	
.	.	.	.	.	.	
Xn-1	$A1_{n-1}$	$A2_{n-1}$	$A3_{n-1}$	$A4_{n-1}$	$A5_{n-1}$	
Xn	$A1_n$	$A2_n$	$A3_n$	$A4_n$	$A5_n$	

2.2. Normalization

Normalization is utilized to provide steady distribution of data sets employed by classification algorithms, to reduce their significance suppressing scale differences of sensor data, and to prevent from performance loses in the stages of learning and classification. Selected attribute set is represented by A=A1, A2, …, An-2, An-1, An, whose normalization formula can be provided as

$$\text{NormA} = \frac{A_i - A_{min}}{A_{max} - A_{min}} \tag{1}$$

where A_i refers to attribute value to be normalized, NormA denotes normalized attribute value, A_{min} indicates minimum attribute value, and A_{max} represents maximum attribute value.

2.3. Training Aggressive/calm Driving using Specified Classification Algorithm

In order to realize training process in the classification algorithms, CAN bus data should be labelled as aggressive or calm subsequently to data acquisition process.

Therefore, driving style can be classified in the end of supervised learning process. Here, we prefer binary labelling in which aggressive driving is indicated by "1", and calm driving is represented by "0". Table II comprises normalized aggressive and calm driving dataset. Figure 3 symbolizes driving style in binary form

Table 2. Normalized aggressive and calm driving datasets

	A1	A2	A3	A4	A5
1	$NormA1_1$	$NormA2_1$	$NormA3_1$	$NormA4_1$	$NrmA5_1$
1	$NormA1_2$	$NormA2_2$	$NormA3_2$	$NormA4_2$	$NrmA5_2$
.	.	.	.	.	.
0	$NormA1_n$	$NormA2_n$	$NormA3_n$	$NormA4_n$	$NrmA5_n$

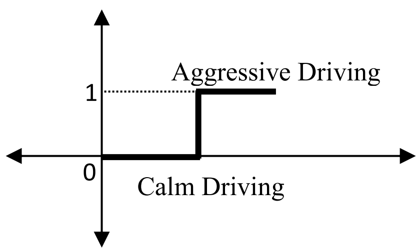


Figure 3. Labelling driving style in binary form

2.3.1. Artificial Neural Network

One of the classification algorithms in the proposed approach is Artificial Neural Network (ANN). Selection of transfer function is a significant stage for ANN. Correlation coefficient is utilized in this selection, which is provided in Equation 2. Correlation coefficient (r) comprising magnitude and direction, varying from -1 to +1, gives the relationship between estimated and actual results. Hence, the relationship between ground truth and estimated result can be evaluated.

$$r_{xy} = \frac{\sum xy - \frac{(\sum x)(\sum y)}{N}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{N}\right]\left[\sum y^2 - \frac{(\sum y)^2}{N}\right]}} \quad (2)$$

where x refers to independent variable, y represents dependent variable, and N indicates number of observations. All the correlation results are plotted in Figure 4, which are Linear Transfer Function (PURELIN), Unipolar Sigmoid Transfer Function (LOGSIG), and Bipolar Sigmoid Transfer Function (TANSIG). Here, the TANSIG being at closest value to +1 appears most stable estimation.

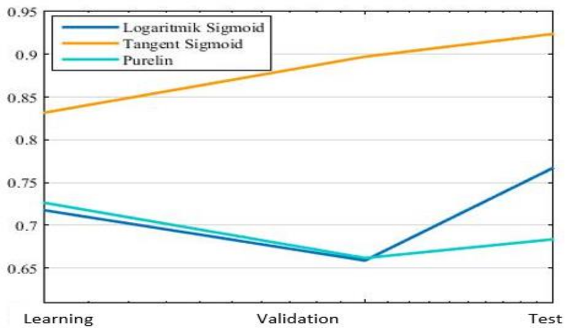


Figure 4. Correlation values of the transfer function

Table 3. Correlation coefficients

	Learning	Validation	Test
	r	r	r
LogSig	0,71744	0,65875	0,76662
TanSig	0,83182	0,89724	0,92370
Purelin	0,72618	0,66166	0,68348

Transfer estimation is given by

$$F(NET) = \frac{(e^{NET} + e^{-NET})}{(e^{NET} - e^{-NET})} \quad (3)$$

considering TANSIG which is selected as transfer function. Designed ANN model is illustrated in Figure 5.

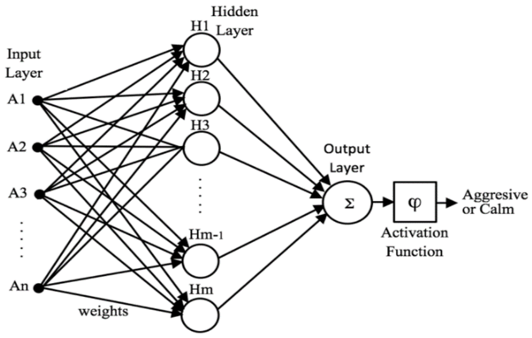


Figure 5. Designated ANN

2.3.2. Naive Bayes

Another classification algorithm which is employed by this study is Naïve Bayesian algorithm. Each of attributes located in this classifier is conditional independent. Trained variable is conditional dependent to these attributes. In the statistical calculations of the Naïve Bayesian algorithm, data sets are firstly determined. Given i=1, …, n data sets can be defined as $x_i = [A1_i, A2_i, A3_i, A4_i, A5_i, \ldots]$. Each of these data sets have m number of classes (cj). Aggressive or calm probability of class j is estimated by

$$P(cj\,|x) = P(x|cj\,) \cdot P(cj\,)/P(x) \quad (4)$$

Independent features are also obtained by

$$P(c_j|x) = \prod_{i=1}^{n} P(x_i|c_j) = P(x_1|c_j) \cdot P(x_2|c_j) \cdot \ldots \cdot P(x_n|c_j) \quad (5)$$

2.3.3. K-Nearest Neighbors

The KNN classification algorithm possesses a memory based structure. For this reason, data to be classified are repeatedly calculated in each iteration. In the algorithm, variables replacing in data sets are represented by a space with n-dimensional as many as sample data numbers, which are as shown in Table I. Each training data indicates a point $\{A1_i, A2_i, A3_i, A4_i, A5_i, \ldots, An_i\}$ in n-dimensional space. Following training stages, new samples referring to test data of the system are added to the nearest neighbor class determining K-number of neighbors through training data. While the classification is proceeded, distances to variables of groups are calculated and one with shortest distance is obtained. This variable set is added to the class. For n-dimensional space, the coordinate plane indicating variable set class to be added is given in Figure 6. There is different distance estimation algorithms in literature. In this study, well-known Euclidian distance measure is performed. Xu in Figure 6 denotes data to be classified. Euclidian distance calculation to find class of data is provided in.

$$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (6)$$

where x_i and y_i represent distance at horizontal and vertical axis

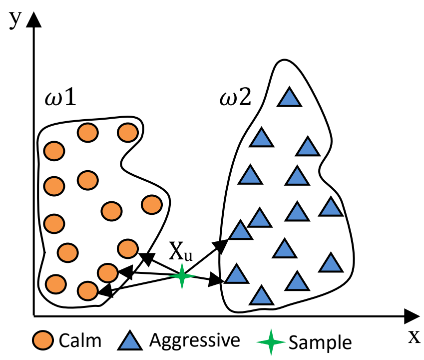


Figure 6. Correlation plane of KNN classification

2.3.4. Support Vector Machine

In this study, SVM is aimed to find hyper plane virtual line separating calm and aggressive driving classes at equal and furthest distance. Dashed lines determine the border line of closest data to the other groups. Thus, the distance between these two groups is divided into two portions through the hyper plane. Class of streaming test data is obtained by border line at equal distance to both class using training data. Hyper plane line discriminating aggressive and calm driving is determined by SVM coordinate system.

2.3.5. C4.5 Algorithm (J48 Algorithm)

One of the algorithms that is preferred in this study is C4.5 or J48 algorithm. Decision trees are utilized for classification in C4.5 algorithm. It can continue classifying even if missing information occurs. Threshold value of each variable set, for example k=1..., n and $A5_k = \{A5_1, A5_2, A5_3, \ldots, A5_n\}$ in Table I, should be determined for classification, and all the values located in variable set should be sorted. The threshold value is obtained by calculating the median value of these sorted data as

$$T_i = \frac{A_k + A_{k+1}}{2} \tag{7}$$

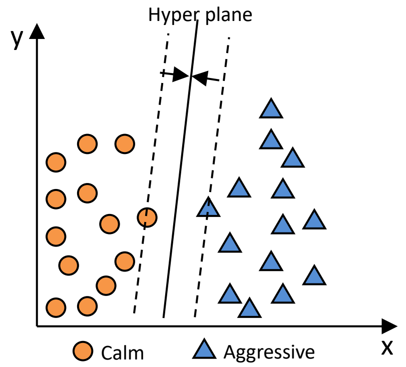


Figure 7. SVM coordinate plane

To find locations of data group in branch of tree, the entropy of each one should be obtained right after getting threshold values. Uncertainty measure for each parameter is thus determined, and these parameters are sorted considering their discrimination level. Entropy values are calculated by

$$H(S) = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{8}$$

where H(S) refers to entropy values of S, n denotes number of messages generated, and p_i indicates probability for generating messages.

2.4. Validation and Performance Evaluation

In order to reveal the relationship between the variables in the dataset, training data which are labelled as aggressive or calm are trained using classification algorithms, and thus aggressive or calm driving classification is performed. Estimation performance for ANN can be given by

$$MSE = \frac{1}{k} \sum_{i=1}^{k} (y_i - \tilde{y_i})^2 \tag{9}$$

where MSE refers to Mean Square Error, k indicates number of data, yi represent value at instant i, and $\tilde{y_i}$ denotes i th predicated values at instant i. Here, the least error means the most accurate estimation. The validation for Naïve Bayes, SVM, KNN, and C4.5 is achieved by recall and precision values. There is a reverse correlation between recall and precision values as shown in Figure 8.

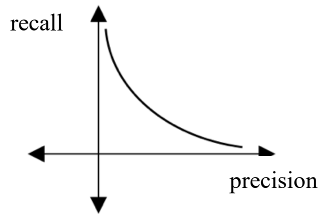


Figure 8. The relation between recall and precision

In some situations, while precision result is high, recall value may be low. For that reason, the result may not be meaningful as shown in Figure 8. Therefore, harmonic average value of arising results is obtained to get F-Score value, which refers to performance value. Table IV includes data classification results in which TP denotes True Positive, FN represents False Negative, FP indicates False Positive, and TN refers to True Negative. Furthermore, Equation 10 denotes accuracy rate estimation, Equation 11 represents error rate estimation, Equation 12 refers to precision estimation, Equation 13 indicates recall estimation, and Equation 14 reveals F-Score estimation (Sokolova at al., 2006).

Table 4. Data classification results (confusion matrix)

	Estimated (No)	*Estimated (Yes)*	
Actual (No)	TN	FP	*TN + FP*
Actual (Yes)	FN	TP	*FN + TP*
	TN + FN	*FP + TP*	

$$\text{Accuracy rate} = \frac{TP+TN}{TP+FP+FN+TN} \tag{10}$$

$$\text{Error rate} = \frac{FP+FN}{TP+FP+FN+TN} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{12}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{13}$$

$$\text{F-Score} = \frac{2 \cdot \text{recall} \cdot \text{precision}}{\text{recall}+\text{precision}} \tag{14}$$

3. Experimental Results

3.1. Data Acquisition via CAN bus

In this study, five different drivers get a specified lap in two times for collecting data as shown in Figure 9. Drivers were asked for driving either aggressive or calm in each trial lap. Therefore, this process provides a table including ground truth data for aggressive and calm driving through same route. The main goal of this study is to estimate driving style of drivers through OBDII and to compare achievements of classification algorithms. Classification process needs characteristically attributes. All the sensor data acquired by OBDII is not distinctive. For this reason, optimum variables are selected to accomplish classification, which are A1, A2, A3, A4, and A5 referring to MAF (grams/sec), Calculated Load (%), RPM, Instant Economy Cost (100 Km/L), and Vehicle Speed (Km/h), respectively. The rest does not have characteristically attributes for discriminating aggressive or calm

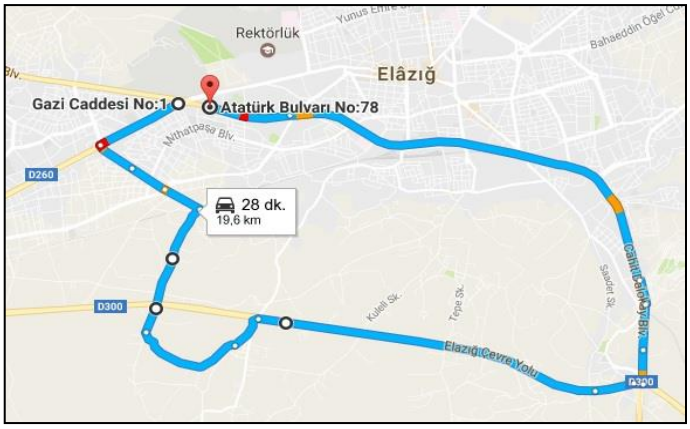


Figure 9. Data collection route

Figure 10. Data acquisition via OBDII

driving. The specified route on which sensor data is collected is shown in Figure 9. Selected sensor data for the specified route is provided in Table V addressing raw training data set. In the Driving Type column of the table, aggressive and calm driving are labelled as “1” and “0”, respectively. The sensor data through CAN bus varies by vehicle brand and series. Selected attributes are common for most of the vehicles. Aggressive dataset consist of data of 281 rows, and calm data set consist of 372 rows. Sample snapshots for data acquisition experiment in vehicle are shown in Figure 10.

Table 5. Raw training dataset

		A1	**A2**	**A3**	**A4**	**A5**
No	**Driving Type**	**MAF**	**Calculated Load**	**RPM**	**Instant Economy**	**Vehicle Speed**
1	1	5,74	96,8	2246	10,5	39,8
2	1	6,11	98,8	2411	10,6	42,9
.	1	.	.	.	.	.
281	1	5,54	87,8	2170	10,5	38,5
282	0	0,87	14,9	861	1	0,6
283	0	0,94	15,7	805	1,1	0,6
.	0	.	.	.	.	.
652	0	0,85	14,9	865	2	1,2
653	0	0,87	14,9	861	1	0,6

A1: MAF (grams/sec), A2: Calculated Load (%), A3: RPM, A4: Instant Economy Coast (100 Km/L), A5: Vehicle Speed (Km/h)

Table 6. Normalized training datasets

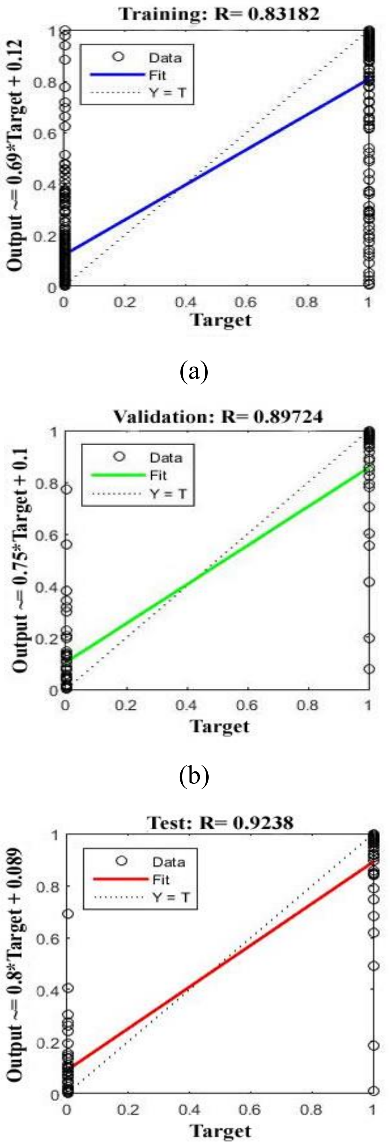
		A1	**A2**	**A3**	**A4**	**A5**
No	**Driving Type**	**MAF**	**Calculated Load**	**RPM**	**Instant Economy**	**Vehicle Speed**
1	1	0,685	0,968	0,399	0,178	0,762
2	1	0,735	0,988	0,444	0,179	0,821
.	1	.	.	.	.	.
281	1	0,659	0,878	0,378	0,178	0,737
282	0	0,034	0,149	0,022	0,016	0,011
283	0	0,044	0,157	0,007	0,018	0,011
.	0	.	.	.	.	.
652	0	0,032	0,149	0,023	0,033	0,023
653	0	0,034	0,149	0,022	0,016	0,011

3.2. Normalization

Collected data should be normalized before training, validation, and test processes. Therefore, relative amplitude differences in sample data are suppressed, and also this process minimizes ill-effect of input data with outliers on the running system operation. The input data are taken into the normalization range between 0 and +1 as given in Equation 1 denoting min-max operation. Normalized data for aggressive and calm driving are given in Table VI.

3.3. Training, validation, and test for ANN

The selection of transfer function to be employed by ANN is achieved, which is given in Equation 2. Correlation performance of transfer functions are obtained as shown in Figure 4. Then, the bipolar sigmoid transfer function (TANSIG) exposing best performance is selected. Number of hidden layer in ANN is determined by some factors including over fitting and performance to be involved in optimum layer. As it can be remembered, the layout of the proposed ANN model is provided in Figure 5. Distribution of training, validation, and test process results for the proposed ANN is shown in Figure 11.



(a)

(b)

(c)

Figure 11. Performance results of ANN, (a) training performance, (b) validation performance, (c) test process performance

3.4. Performance measure of Naïve Bayes, KNN, SVM, C4.5 and ANN

Confusion matrix for the investigated algorithms is provided in Table VII, considering 94% training data and 6% test data. The results in confusion matrix is substituted by Equation 10, 11, 12, 13, and 14; then performance results are obtained as given in Table VIII including accuracy rate, error rate, precision, recall, and F-Score.

Table 7. Confusion matrix

	Performance Results			
The Algorithms	TN	FN	FP	TP
Naive Bayes	17	2	1	20
KNN	15	3	3	19
SVM	17	4	1	18
C4.5 (J48)	15	3	3	19
ANN	16	3	2	19

TP: True Positive FN: False Negative
FP: False Positive TN: True Negative

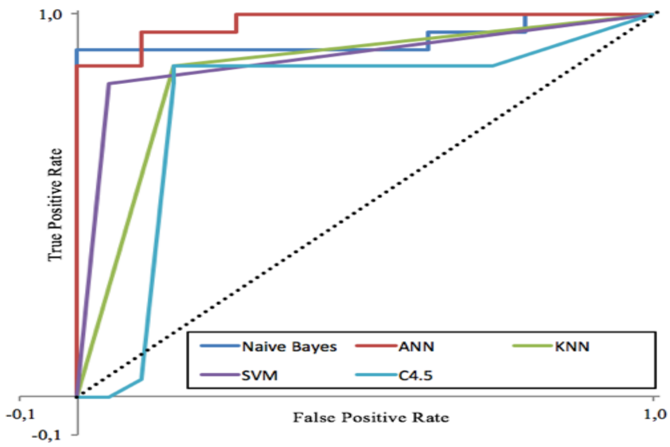


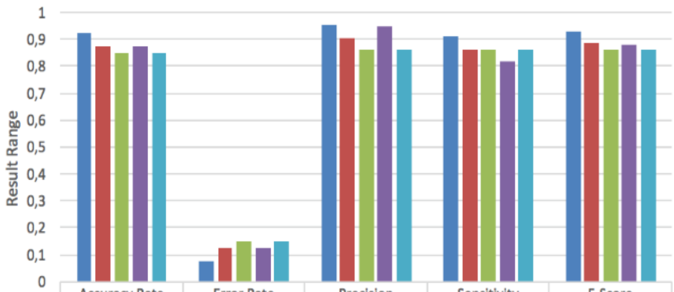
Figure 12. ROC curves indicating performance of the algorithms

Table 8. Performance results the algorithms

	Performance Measures				
Algorithms	**Accuracy Rate**	**Error rate**	**Precision**	**Recall**	**F-Score**
Naive Bayes	0,925	0,075	0,95	0,909	0,930
KNN	0,850	0,150	0,86	0,864	0,864
SVM	0,875	0,125	0,94	0,818	0,878
C4.5 (J48)	0,850	0,150	0,86	0,864	0,864
ANN	0,875	0,125	0,90	0,864	0,884

3.5. Validation and Performance Evaluation Process

The validation process has been carried out using new aggressive and calm driving data which is different from previous training data. In the acquisition stage, raw test data addressing to calm and aggressive driving have been recorded. Then, normalization process is applied for them. And then, aforementioned algorithms have been experimented by test data.



	Accuracy Rate	Error Rate	Precision	Sensitivity	F-Score
Naive	0,925	0,075	0,952	0,909	0,93
ANN	0,875	0,125	0,905	0,864	0,884
KNN	0,85	0,15	0,864	0,864	0,864
SVM	0,875	0,125	0,947	0,818	0,878
C4.5	0,85	0,15	0,864	0,864	0,864

Figure 13. Performance results of the algorithms in bar representations

Performance results are provided in Table VII. By means of recall (sensitivity) and precision values, we can obviously perform a logical interpretation between both indicators being inverse parabolic relationship. Therefore, we have focused on interpreting of F-Score and harmonic average values obtained by Equation 14. Further, accuracy rate gives a valuable clue to have an idea for performance of existing algorithms. On the other hand, in Table VIII, accuracy rate and F-Score values of the algorithms are close to each other.

ROC curve referring to test results of the classification algorithms are shown in Figure 12. Also, Table VII is converted to bar representation for visually clarifying the issue as provided in Figure 13. Accuracy rate, F-Score, precision, and recall (sensitivity) scores of Naïve Bayes algorithm are better than those of the others, and ROC curve validates this situation. Hence, their error rates are less than those of the rest.

4. Conclusions and Recommendations

In the present study, we have compared the performance results of specified algorithms for aggressive and calm driving by means of limited internal sensors. According to the F-Score rate of 93%, Naïve Bayes is estimated as best performance value. The F-Score performance rates of Artificial Neural Network, Support Vector Machine, K-Nearest Neighbors, and C4.5 (J48) algorithms are calculated as 88%, 87%, 86%, and 86% in ascending order. The difference between both algorithms showing best and least performance is 7%, which is the meaningful result in specified trials. In the next study, we are planning to deploy embedded software estimating driving style through driving mood involving in Fuzzy Logic approach for aggressive and calm driving mood.

References

https://www.safemotorist.com/articles/road_rage.aspx (Last access 06.12.2020)

https://www.iii.org/fact-statistic/facts-statistics-aggressive-driving (06.12.2020)

https://aaafoundation.org/2015-traffic-safety-culture-index/ (06.12.2020)

http://www.tuik.gov.tr/ (06.12.2020)

Kumtepe, Ö., Akar, G. B., & Yüncü, E. (2015). On Vehicle Aggressive Driving Behavior Detection Using Visual Information. In *2015 Signal Processing and Communications Applications Conference (SIU)* (pp. 795-798).

Kumtepe, Ö., Yüncü, E., & Akar, G. B. (2016). A Multimodal Approach for Aggressive Driving Detection. In *2016 Signal Processing and Communication Application Conference* (pp. 729-732).

Johnson, D. A., & Trivedi, M. M. (2011, October). Driving style recognition using a smartphone as a sensor platform. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)* (pp. 1609-1615). IEEE.

Eren, H., Makinist, S., Akin, E., & Yilmaz, A. (2012, June). Estimating driving behavior by a smartphone. In *2012 IEEE Intelligent Vehicles Symposium* (pp. 234-239). IEEE.

Bergasa, L. M., Almería, D., Almazán, J., Yebes, J. J., & Arroyo, R. (2014, June). Drivesafe: An app for alerting inattentive drivers and scoring driving behaviors. In *2014 IEEE Intelligent Vehicles symposium proceedings* (pp. 240-245). IEEE.

Koh, D. W., & Kang, H. B. (2015, June). Smartphone-based modeling and detection of aggressiveness reactions in senior drivers. In *2015 IEEE Intelligent Vehicles Symposium (IV)* (pp. 12-17). IEEE.

Li, F., Zhang, H., Che, H., & Qiu, X. (2016, November). Dangerous driving behavior detection using smartphone sensors. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1902-1907). IEEE.

Roider, O., Wegener, S., Stark, J., Judmaier, P., Michelberger, F., & Barberi, A. (2019). Merging Virtual World with Real-Life Behavior: A Concept for a Smartphone App to Influence Young People's Travel Behavior. *Transportation research record*, *2673*(4), 241-250.

Imkamon, T., Saensom, P., Tangamchit, P., & Pongpaibool, P. (2008, May). Detection of hazardous driving behavior using fuzzy logic. In *2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology* (Vol. 2, pp. 657-660). IEEE.

Wu, B. F., Chen, Y. H., & Yeh, C. H. (2012, November). Fuzzy logic based driving behavior monitoring using hidden markov models. In *2012 12th International conference on ITS telecommunications* (pp. 447-451). IEEE.

Songkroh, A., Fooprateepsiri, R., & Lilakiataskun, W. (2014, June). An intelligent risk detection from driving behavior based on BPNN and Fuzzy Logic combination. In *2014 IEEE/ACIS 13th International Conference on Computer and Information Science (ICIS)* (pp. 105-110). IEEE.

Fazio, P., Santamaria, A. F., De Rango, F., Tropea, M., & Serianni, A. (2016, May). A new application for analyzing driving behaviour and environment characterization in transportation systems based on a fuzzy logic approach. In *Unmanned Systems Technology XVIII* (Vol. 9837, p. 983707). International Society for Optics and Photonics.

Arefnezhad, S., Samiee, S., Eichberger, A., & Nahvi, A. (2019). Driver drowsiness detection based on steering wheel data applying adaptive neuro-fuzzy feature selection. *Sensors*, *19*(4), 943.

Vesselenyi, T., Rus, A., Mitran, T., Moca, S., & Lehel, C. (2019, October). Fuzzy Decision Algorithm for Driver Drowsiness Detection. In *SIAR International Congress of Automotive and Transport Engineering: Science and Management of Automotive and Transportation Engineering* (pp. 458-467). Springer, Cham.

Vaitkus, V., Lengvenis, P., & Žylius, G. (2014, September). Driving style classification using long-term accelerometer information. In *2014 19th International Conference on Methods and Models in Automation and Robotics (MMAR)* (pp. 641-644). IEEE.

Li, Y., Miyajima, C., Kitaoka, N., & Takeda, K. (2014, October). Measuring aggressive driving behavior using signals from drive recorders. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)* (pp. 1886-1887). IEEE.

Vignali, V., Bichicchi, A., Simone, A., Lantieri, C., Dondi, G., & Costa, M. (2019). Road sign vision and driver behaviour in work zones. *Transportation research part F: traffic psychology and behaviour*, *60*, 474-484.

de Naurois, C. J., Bourdin, C., Stratulat, A., Diaz, E., & Vercher, J. L. (2019). Detection and prediction of driver drowsiness using artificial neural network models. *Accident Analysis & Prevention*, *126*, 95-104.

Doshi, A., & Trivedi, M. M. (2010, June). Examining the impact of driving style on the predictability and responsiveness of the driver: Real-world and simulator analysis. In *2010 IEEE Intelligent Vehicles Symposium* (pp. 232-237). IEEE.

Gregoriades, A., Florides, C., Lesta, V. P., & Pampaka, M. (2013, October). Driver behaviour analysis through simulation. In *2013 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 3681-3686). IEEE.

Shirazi, M. M., & Rad, A. B. (2014). Detection of intoxicated drivers using online system identification of steering behavior. *IEEE Transactions on Intelligent Transportation Systems*, *15*(4), 1738-1747.

Keklikoglou, A., Fitzpatrick, C. D., & Knodler, M. A. (2018). Investigation of time and speed perception using a driving simulator. *Transportation research record*, *2672*(37), 132-140.

Dia, H., & Panwai, S. (2015, December). Impact of Driving Behaviour on Emissions and Road Network Performance. In *2015 IEEE International Conference on Data Science and Data Intensive Systems* (pp. 355-361). IEEE.

Sun, J., Zhen, A., Li, C., Zhang, M., & Hu, X. (2016, October). A vehicles' CO 2 emission monitoring platform combined with driver's driving behavior. In *2016 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)* (pp. 1-2). IEEE.

Stogios, C., Kasraian, D., Roorda, M. J., & Hatzopoulou, M. (2019). Simulating impacts of automated driving behavior and traffic conditions on vehicle emissions. *Transportation Research Part D: Transport and Environment*, *76*, 176-192.

Faria, M. V., Duarte, G. O., Varella, R. A., Farias, T. L., & Baptista, P. C. (2019). Driving for decarbonization: assessing the energy, environmental, and economic benefits of less aggressive driving in Lisbon, Portugal. *Energy Research & Social Science*, *47*, 113-127.

Karaduman, O., Eren, H., Kurum, H., & Celenk, M. (2013, December). An effective variable selection algorithm for Aggressive/Calm Driving detection via CAN bus. In *2013 International Conference on Connected Vehicles and Expo (ICCVE)* (pp. 586-591). IEEE.

Taylor, A., Japkowicz, N., & Leblanc, S. (2015, December). Frequency-based anomaly detection for the automotive CAN bus. In *2015 World Congress on Industrial Control Systems Security (WCICSS)* (pp. 45-49). IEEE.

Fugiglando, U., Massaro, E., Santi, P., Milardo, S., Abida, K., Stahlmann, R., ... & Ratti, C. (2018). Driving behavior analysis through CAN bus data in an uncontrolled environment. *IEEE Transactions on Intelligent Transportation Systems*, *20*(2), 737-748.

Lokman, S. F., Othman, A. T., Bakar, M. H. A., & Musa, S. (2019, July). The Impact of Different Feature Scaling Methods on Intrusion Detection for in-Vehicle Controller Area Network (CAN). *In International Conference on Advances in Cyber Security* (pp. 195-205). Springer, Singapore.

Le, Q., Jiang, K., & Zhang, F. (2020). Design of automatic detection system for vehicle networking communication abnormal data based on CAN bus. *International Journal of Information and Communication Technology*, *16*(2), 123-139.

Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006, December). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. *In Australasian joint conference on artificial intelligence* (pp. 1015-1021). Springer, Berlin, Heidelberg.