International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Driving Style Recognition for Intelligent Vehicle using Unsupervised Clustering Algorithms

Abhishek Dixit[1], Manish Jain[2]

Submitted: 10/02/2024 **Revised**: 16/03/2024 **Accepted**: 20/03/2024

Abstract: The manner in which a driver operates their vehicles has a significant impact on both energy management and driving safety. Moreover, it is a crucial factor in the advancement of driver assistance systems (ADAS), which aim to increase the level of vehicle automation. As a result, numerous research and development initiatives have been undertaken to identify and classify driving styles. In this study, we have used principal component analysis for feature reduction and K means clustering algorithm for driving style identification of the vehicle. To evaluate the performance of the proposed approach, it was tested using vehicle trajectory data from the Next Generation Simulation (NGSIM) project, specifically the datasets collected on US Highway 101 and I-80. The proposed approach introduces a novel method that enhances efficiency and accuracy, offering a significant advancement in addressing complex challenges within its respective domain.

Keywords: *Driving style, driving style recognition, intelligent vehicle control, machine learning, connected vehicle*.

1. Introduction

Driving style is defined as the way in which a driver usually drives the vehicle and it plays a significant role in ensuring the safety of the driver. Risk of encountering an accident largely depends upon the driving style of an individual. Thus, by determining the driving style, the chances of road accidents can be reduced to some extent. In order to recognize the driving pattern, trajectory data needs to be analyzed carefully along with the machine learning algorithms to predict the state of the individual [1]. Cameras can be installed at the places to record the trajectory details along with the speed, acceleration and other required information of the vehicle.

The kind of driving style plays a very crucial guide and an important measurement of a driver's ability to perform and drive in a protective manner. The style of driving depends on various factors such as emotion, motor, velocity, acceleration jerk and these may change over space and time and using this we can classify them into different styles that will help us to analyze the situations in a much deeper way.

Safe and efficient operation of autonomous vehicles requires accurate prediction of the future trajectories of surrounding vehicles. However, uncertainties in real-world driving scenarios make trajectory forecasting challenging. Other vehicles may exhibit varying driving styles and unpredictably deviate from expected behaviors [2]. To enable robust decision-making, autonomous vehicles need to reason about likely trajectory distributions amidst uncertainties.

To understand further we have done a driving style profiling, it is a type of process in which we collect driver’s data for different data points and try to classify it depending on certain parameters.

Some of the essential factors in user acceptance of autonomous vehicles are safety, reliability, and comfort. We can determine the different style of driving by modeling different parameters of its motion plane trajectory. In recent years, research in autonomous driving has surged, driven by technological advancements and improvements in user experience. Particularly, the significant strides in the fields of AI and machine learning have been a leap of faith for researchers, reducing the need for human interventions and enhancing the overall interaction between customers and autonomous vehicles.

Clustering algorithms play a very crucial role in determining the driving style of a person, for autonomous vehicles, recognizing the driving patterns of nearby human-operated cars is an important capability for making safe yet efficient driving decisions, particularly in environments with a mix of automated and human-driven vehicles. Collecting driving data over different time periods could help characterize how human driver behavior varies based on parameters like time of day. Cluster analysis of such multivariate driving data could reveal insights into distinct human driving styles, which would assist autonomous vehicles in responding appropriately to each style.. The driving data we took is unlabeled and only a period of driving data can help predict the drivers driving style. It becomes a little challenging to predict style using

[1] JECRC University, Jaipur – 303905, India
ORCID ID : 0000-0003-1294-9654
[2] JECRC University, Jaipur – 303905, India
ORCID ID : 0000-0002-3535-0030
** Corresponding Author Email: abhishekdixit2606@gmail.com*

unlabeled data so we have proposed an unsupervised learning approach to accurately classify the style of driving for autonomous drivers.

Nowadays the industry itself has taken a step forward to make autonomous driving better by providing good amount of feedback towards safe driving practices and that has helped in reducing number of accidents and increased road safety to a much greater extent. The kind of driving style plays a very crucial role and an important measurement of a driver's ability to perform and drive in a protective manner. The style of driving depends on various factors such as emotion, motor, velocity, acceleration jerk and these may change over space and time and using this we can classify them into different styles that will help us to analyze the situations in a much deeper way. To understand further we have done a driving style profiling, it is a type of process in which we collect driver’s data for different data points and try to classify it depending on certain parameters.

This paper proposes the recognition model for the driving behaviour which is based on the vehicle trajectory data. The model proposed in this paper will basically categorize the driver into three main categories as aggressive driver, moderate driver and traditional driver surrounding the autonomous vehicle.

2. Related Work

Categorizing the driving style of vehicles surrounding an autonomous vehicle is a crucial aspect of developing a safe and efficient autonomous system. Unsupervised clustering algorithms have been widely used to categorize the driving styles of surrounding vehicles, due to their ability to identify patterns in large datasets without prior knowledge or labeling. This literature review summarizes recent research in this field, focusing on the use of unsupervised clustering algorithms for categorizing the driving style of surrounding vehicles.

[3]In a study by Lee et al. (2018), the authors used a k-means clustering algorithm to categorize the driving styles of surrounding vehicles based on their speed, acceleration, and lane-keeping behavior. The results showed that the k-means algorithm was able to accurately categorize the driving styles into three distinct groups: aggressive, normal, and cautious.

[4]Another study by Chen et al. (2019) used a hierarchical clustering algorithm to categorize the driving styles of surrounding vehicles based on their speed, acceleration, and deceleration patterns. The results showed that the hierarchical clustering algorithm was able to identify four distinct driving styles: aggressive, normal, cautious, and hesitant.

[5]In a study by Wang et al. (2020), the authors used a Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to categorize the driving styles of surrounding vehicles based on their speed, acceleration, and lateral position relative to the autonomous vehicle. The results showed that the DBSCAN algorithm was able to accurately categorize the driving styles into three groups: aggressive, normal, and cautious.

Additionally, there are studies that combined multiple clustering algorithms to improve the accuracy of driving style categorization. For example, a study by Zhang et al. (2021) used a hybrid approach combining k-means and Gaussian Mixture Model (GMM) algorithms to categorize the driving styles of surrounding vehicles. The results showed that the hybrid approach was able to accurately categorize the driving styles into four groups: aggressive, normal, cautious, and unpredictable [6].

In conclusion, unsupervised clustering algorithms have been widely used in recent research for categorizing the driving style of vehicles surrounding an autonomous vehicle. The most commonly used algorithms include k-means, hierarchical clustering, DBSCAN, and hybrid approaches combining multiple algorithms. These algorithms have been shown to be effective in categorizing the driving styles into distinct groups, such as aggressive, normal, cautious, and hesitant.

The method adopted in this paper introduces a novel method that enhances proficiency and precision, offering a critical progression in tending to complex challenges within its respective domain.

3. Methodology

In this research, the dataset or vehicle trajectory pairs are derived from the data obtained from the I-80 and US-101 freeways. The NGSIM vehicle trajectory data, collected from a distinct region and time, offer insights into both congested and moderate traffic scenarios. The preprocessing of data is a crucial step in the machine learning process. This step involves finding and filling in any missing values, removing the outliers, ensuring that the correct data set is used and extracting relevant features [7]. The format of the data set is essential for analysis. Feature selection is done through the correlation plots and it has been observed that the data is randomly distributed in all the comparison plots which imply that all features are independent and make effect the output label. The data collected during this stage will be input into the Google colab platform as python programming to obtain the desired output. Various ML algorithms are applied to the pre-processed dataset and results are evaluated. The figure 1 mention the method adopted for driving style prediction.

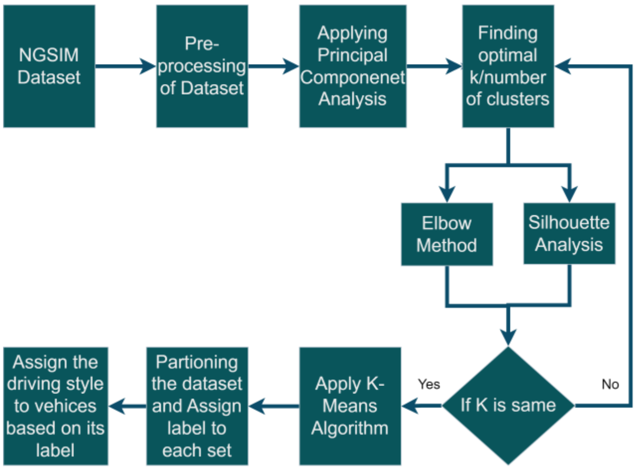

Fig. 1. Method adopted for driving style prediction

3.1. NGSIM Data Preprocessing

The NGSIM dataset is meticulously labeled with vehicle and frame identities, making it conducive for analyzing car-following behaviors [8]. The study focuses on assembling leading-following pairs from various vehicles and locations.

An inherent challenge with the raw NGSIM data lies in the presence of notable data errors. For instance, the natural trajectory in NGSIM may result in a vehicle generating a collision path, a scenario that deviates from real-world observations [9]. To address this issue, a data pre-processing step is implemented in this section to rectify the raw dataset. The data cleaning procedure outlined in [10] serves as a basis, with a focus on interpolating impossible kinematics such as abrupt zero speed, unfeasibly high acceleration values, or sudden zero headway space using neighboring data [11]. Subsequently, points exhibiting significant abnormality like those exceeding 30 times the mean value or identified as unusual cases are replaced with values derived from the mean.

3.2. Applying Principal Component Analysis (PCA) on the dataset

Feature indicators undergo dimensionality reduction to simplify the intricate relationship within the data and identify key features characterizing driving styles. Principal component analysis (PCA), a commonly employed method for dimension reduction, generates new components through linear combinations based on the correlation matrix of original variables. In PCA, a criterion is set to ensure that the cumulative contribution rate of selected components surpasses 60%, serving as a measure of the retained original characteristic information [12]. Upon processing trajectory data, PCA calculations were executed, and the resulting cumulative contribution rate is depicted in the figure 2.

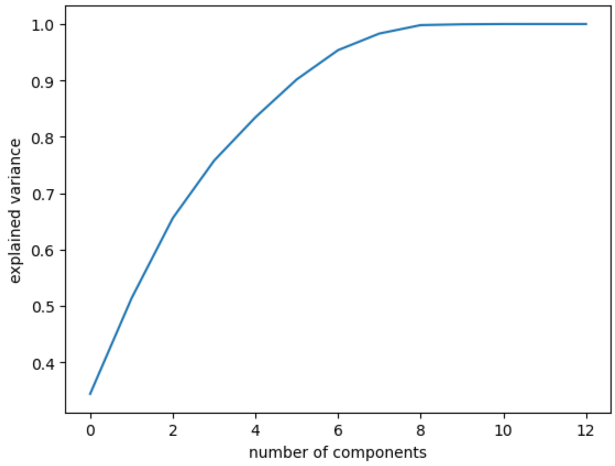

Fig. 2. Cumulative contribution rate of principal component

With the extraction of the first two principal components, the cumulative contribution rate exceeded 60%, indicating that these two characteristic indexes effectively represent the majority of information related to merging vehicles [13] as shown in the figure 3.

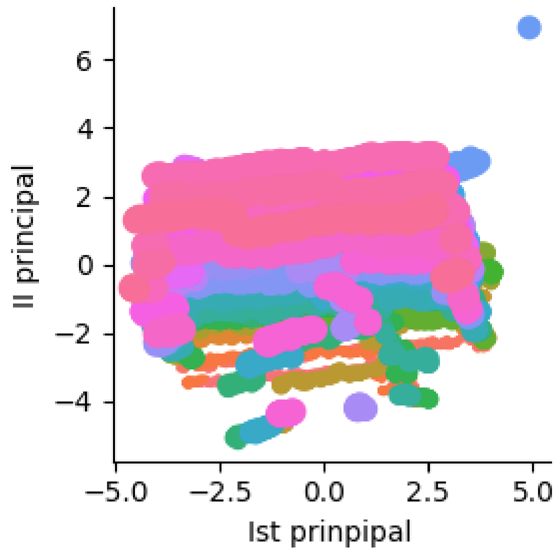

Fig. 3. Cumulative contribution rate of principal component

3.3. Finding Clustering category number using Elbow method & Silhouette analysis

Unsupervised classification in the form of clustering analysis is employed for studying driving styles, with the K-means clustering algorithm being a widely utilized approach. This algorithm relies on distance metrics, where the proximity between two objects indicates their similarity. The fundamental concept involves iteratively categorizing data sets into distinct groups [14]. In this study, the K-means algorithm was applied to classify merging vehicle data based on the distances between various samples and clustering centers. The initial step involved determining the optimal number of clustering categories, denoted as the K value [15]. The classification effectiveness of different K values was evaluated through the "Elbow rule" and Silhouette value calculation for data clustering. The "Elbow rule" assesses the sum of square errors (SSE) within the cluster, representing the sum of squared distances from each point in the cluster to its designated center for different K values [16]. Ideally, a smaller SSE signifies better clustering effectiveness. The SSE value typically converges towards a minimum, and the "Elbow rule" focuses on identifying the turning point in the curve [17].

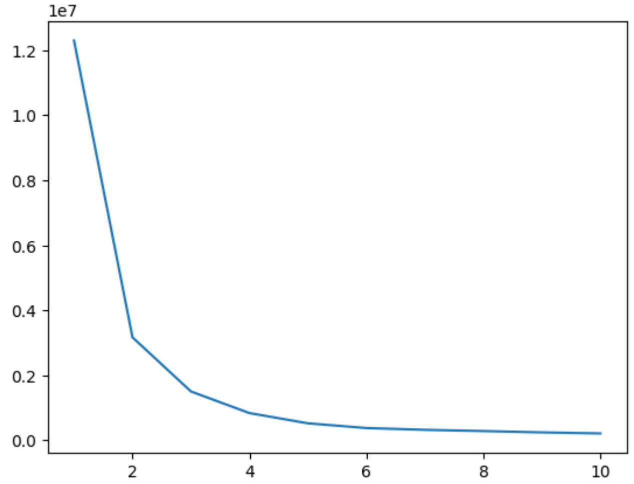

Fig. 4. : Finding clustering category number using Elbow method

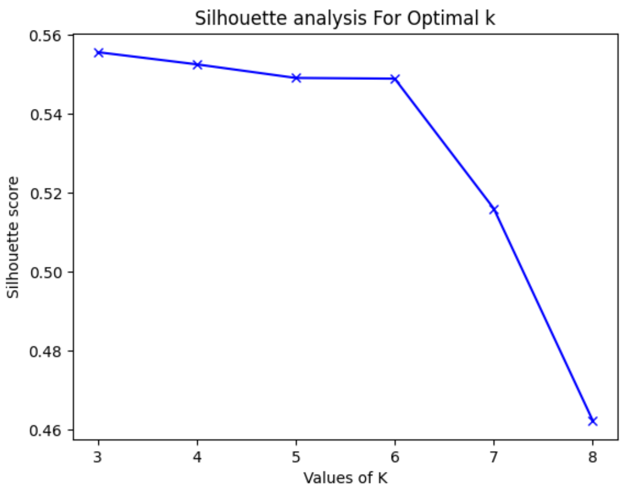

Fig. 5. Finding clustering category number using silhouette analysis

The calculated results depicting the computation outcomes for the SSE value and silhouette coefficient are presented in Figures 5 and 6, respectively. Considering the insights gained from these two indicators, the optimal cluster was determined to be K=3.

3.4. Applying K-means Algorithm

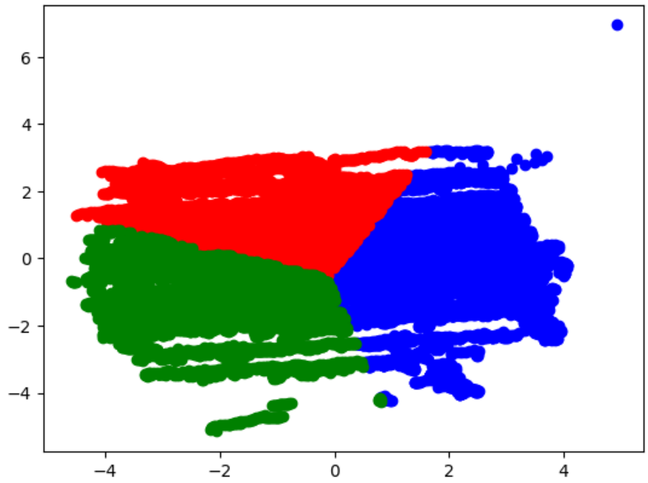

Fig. 6. K-means clustering algorithm

The clustering outcomes, as illustrated in the figure 6, reveal distinct driving styles based on characteristic index values. The classification delineates three types: 0 signifies aggressive driving, 1 denotes a moderate driving style, and 2 represents a traditional driving style.

3.5. Partitioning the dataset and assign labels

We have obtained 3 clusters after applying k means algorithm. On this basis, the labels (0.1.2) are assigned to each value of the dataset.

3.6. Assign the driving style to the vehicle based on label

Specific label is assigned in the previous step to each value of the dataset. The type of vehicle is defined as the maximum number of times it follows the pattern or lie in the particular label. The maximum time a vehicle remain in a particular label or the maximum number of frames of a particular vehicle found in particular label, that label is assigned to that vehicle[15]. The process is shown in the table 1.

4. Results & Analysis

The vehicles are categorized in to three driving styles named as Aggressive, moderate and traditional. From Table 2 it can be seen that the defined aggressive vehicles have faster speed, higher acceleration and jerk, smaller space headway, and lower time headway than the other two groups based on the K-means classification, which shows aggressive drivers may tend to maintain a smaller space with front vehicles with faster speed. Aggressive vehicle also have less time headway compared with others two. The result of moderate vehicle lies in between the aggressive and traditional category when compared on the same parameters like velocity, acceleration, jerk, space and time headway.

Table. 1. Assignment of the driving style to all the vehicles based on maximum labels

Vehicle ID	Total Frames	Label 0	Label 1	Label 2	Max. frame
2	437	35	240	162	1
4	351	63	125	163	2
5	452	44	245	163	1
6	357	82	112	163	2
20	414	68	183	163	1
25	436	97	176	163	1
26	438	74	201	163	1
27	432	107	162	163	2
62	431	165	103	163	0
63	305	177	0	128	0
64	414	161	90	163	2
67	409	167	79	163	0
69	504	258	83	163	0
77	494	273	58	163	0
78	473	242	68	163	0
79	432	209	60	163	0

Table. 2. Comparison of aggressive, moderate and traditional drivers

Driving Style		velocity	Acceleration	Jerk	Space Headway	Time Headway
Aggressive	**mean**	49.08	0.57	5.62E-17	60.82	1.31
	min	24.40	-11.20	-213.7	0.00	0.00
	max	95.30	11.20	224	257.37	5.60
Moderate	**mean**	48.60	0.51	-1.68E-17	82.53	1.73
	min	27.78	-11.20	-195.2	0.00	0.00
	max	77.47	11.20	224	329.31	6.81
Traditional	**mean**	45.55	0.39	-6.29E-17	85.85	3.06
	min	0.00	-11.20	-224	0.00	0.00
	max	75.28	11.20	224	408.24	11.99

Figure 7 depict the comparison graph of aggressive, moderate & traditional vehicle on different feature like velocity, acceleration, jerk, space and time headway. The mean value is taken in to the consideration for comparison among all the categories. For aggressive vehicle, the velocity, acceleration, jerk is higher while space and time Headway is smaller as compared among all the categories of the vehicle i.e. moderate and traditional. All the value of moderate vehicle lie in between aggressive and traditional while the values obtained for the traditional vehicle is inverse of the aggressive vehicles.

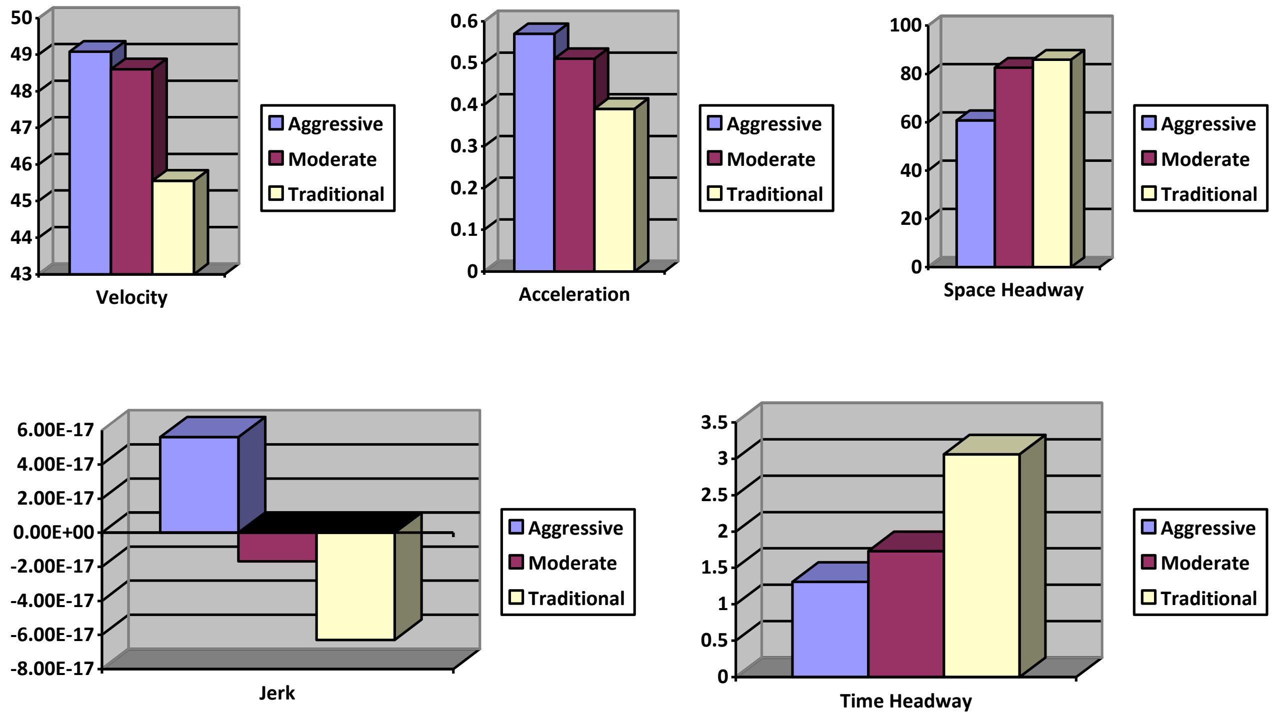

Fig. 7. Comparison graph of aggressive, moderate, traditional vehicle on different feature like velocity, acceleration, jerk, space and time headway

5. Conclusion

This research aimed to categorize vehicle driving styles by analyzing trajectory data. Principal Component Analysis (PCA) is used to streamline characteristic indexes, reducing them to two components that represented all features. The "Elbow rule" and Silhouette methods were utilized to identify the optimal number of categories, and the K-means clustering algorithm was applied to cluster vehicle driving styles. Finally, the driving styles were divided into aggressive, moderate and traditional styles. The study revealed that aggressive driving is characterized by higher velocity, acceleration, and jerk, along with lower space and time headway. The proposed unlabeled data based driving style classification method will be extended to other driving scenarios and new dataset.

References

[1] D. F. Xie, Z. Z. Fang, B. Jia, et al., “A data-driven lane-changing model based on deep learning,” Transportation Research Part C:Emerging Technologies, vol. 106, pp. 41-60, Jul. 2019.

[2] J. Zhang, G. Zhen, H. Jia and H. Wei, "Trajectory Prediction of Surrounding Vehicles for Autonomous Vehicle Using POMDP Method," 2022 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), Nanjing, China, 2022.

[3] J. Lee et al., "Categorizing Driving Styles of Surrounding Vehicles Using K-Means Clustering," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 5, pp. 1548-1557, May 2018.

[4] Y. Chen et al., "Hierarchical Clustering for Categorizing Driving Styles Based on Speed, Acceleration, and Deceleration Patterns," IEEE Transactions on Vehicular Technology, vol. 68, no. 8, pp. 7577-7588, Aug. 2019.

[5] Q. Wang et al., "Categorization of Driving Styles Using DBSCAN Algorithm Based on Speed, Acceleration, and Lateral Position," IEEE Access, vol. 8, pp. 97879-97890, May 2020.

[6] H. Zhang et al., "Hybrid Clustering Approach for Driving Style Categorization Using K-Means and Gaussian Mixture Model," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 3, pp. 1543-1552, Mar. 2021.

[7] T. Xu, X. Zhang and Y. K. Zhang, “Safety Orientation Classification of truck drivers based on AdaBoost algorithm,” Journal of Safetyand Environment, vol. 19, pp. 1273-1281, Aug. 2019.

[8] H. Jin and M. Lv, “Research on Driving Style in starting Condition based on Improved Fisher

criterion,” Transactions of Beijing Institute of Technology, vol. 40, pp. 262-266, Mar. 2020.

[9] Punzo, Vincenzo, Maria Teresa Borzacchiello, and BiagioCiuffo. "On the assessment of vehicle trajectory data accuracy and application to the Next Generation Simulation (NGSIM) program data." Transportation Research Part C: Emerging Technologies 19.6 (2011): 1243-1262.

[10] Coifman, Benjamin, and Lizhe Li. "A critical evaluation of the Next Generation Simulation (NGSIM) vehicle trajectory dataset." Transportation Research Part B: Methodological 105 (2017): 362-377.

[11] Y. Xing, C. Lv and D. Cao, "Personalized Vehicle Trajectory Prediction Based on Joint Time-Series Modeling for Connected Vehicles," in IEEE Transactions on Vehicular Technology, vol. 69, no. 2, pp. 1341-1352, Feb. 2020

[12] Iacono, Teresa, Bould, Emma, BeadleBrown,”An exploration of communication within active support for adults with high and low support needs”. Journal of Applied Research in Intellectual Disabilities, (2019) 32 (1). pp. 61-70.

[13] S. Wen, X. G. Shahd Omar, X. Jin and Z. He, "Analysis of Vehicle Driving Styles at Freeway Merging Areas Using Trajectory Data," 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), Macau, China, 2022, pp. 3652-3656

[14] S. R. Jambula, S. C. Mathi, S. N. Polu, K. G. O. JM, V. P. Mateti and S. Yadav, "Enhanced UAV with Image-Driven Concrete Crack Detection," 2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2023, pp. 1730-1737

[15] M. P. Kiyindou, S. E. Sunday and Z. Hong, "Enhancing the Evaluation Performance of Convolutional Neural Networks-Based Vehicle Classification Systems," 2023 International Conference on the Cognitive Computing and Complex Data (ICCD), Huaian, China, 2023, pp. 68-72.

[16] J. H. Yang, D. J. Kim and C. C. Chung, "Lane Change Intention Inference of Surrounding Vehicle: Comparative Study on Relevance Vector Machine (RVM) and Support Vector Machine (SVM)," 2021 21st International Conference on Control, Automation and Systems (ICCAS), Jeju, Korea, Republic of, 2021, pp. 1580-1585

[17] S. Liu, F. Ren, G. Yang, D. He and Y. Liu, "Employing Deep Unsupervised Learning Method to Identify Testing Scenarios for Automated Vehicles," 2023 10th International Conference on Dependable Systems and Their Applications (DSA), Tokyo, Japan, 2023, pp. 628-635,

[18] A. Adnan, G. M. Mahbubur Rahman, M. M. Hossain, M. S. Mim and M. K. Rahman, "A Deep Learning Based Autonomous Electric Vehicle on Unstructured Road Conditions," 2022 IEEE 12th Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 2022, pp. 105-110.