Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 31 No. 6s (2024)

Trajectory Data Driven Driving Style Recognition for Autonomous Vehicles Using Unsupervised Clustering

Abhishek Dixit[1]*, Manish Jain[2]
[1,2]Department of Computer Science & Engineering, JECRC University, Jaipur, Rajasthan, India
E-mail: [1]abhishekdixit2606@gmail.com, [2]manish.jain@jecrcu.edu.in

Article History:
Received: *29-05-2024*
Revised: *30-06-2024*
Accepted: *20-07-2024*

Abstract:
Introduction: This research focuses on enhancing the understanding and classification of vehicle driving styles by analyzing extensive trajectory data. Recognizing and categorizing different driving behaviours is crucial, particularly for the development of autonomous vehicles, which must predict and respond to the diverse actions of human drivers. Understanding these driving styles is essential for improving road safety and ensuring that autonomous systems can navigate mixed traffic environments effectively.

Objectives: The primary objective of this study is to classify vehicle driving styles into distinct categories like aggressive, moderate, and traditional by leveraging unsupervised learning techniques. This classification aims to improve the predictive capabilities of autonomous vehicles and enhance overall road safety by providing a more nuanced understanding of driving behaviours.

Methods: The study begins by applying Principal Component Analysis (PCA) to simplify complex trajectory data, reducing multiple characteristic indexes into two principal components that encapsulate the most significant features related to driving behaviour. To determine the optimal number of driving style categories, the "Elbow rule" and Silhouette analysis are employed, followed by the application of the K-means clustering algorithm. This approach allows for the effective grouping of driving styles based on the processed data.

Results: The analysis identified three distinct driving styles: aggressive, moderate, and traditional. Aggressive driving is characterized by higher velocities, greater acceleration, and increased jerk, along with shorter space and time headways. Traditional driving styles exhibit more conservative behaviours, with lower speeds, reduced acceleration, and greater following distances. Moderate driving styles lie between these two extremes, reflecting a balanced approach in terms of speed, acceleration, and headway distances.

Conclusions: The findings of this study have significant implications for the development and operation of autonomous vehicles. By accurately classifying driving styles, autonomous systems can better anticipate and react to the behavior of surrounding vehicles, thereby enhancing safety in mixed traffic environments. The research also demonstrates the potential for extending the proposed unsupervised learning approach to other driving scenarios and datasets, offering a scalable solution for ongoing advancements in intelligent transportation systems.

Keywords: Driving style, autonomous vehicle, clustering, driving style recognition, intelligent vehicle, machine learning

1. Introduction

The way a person drives a vehicle significantly impacts their risk of accidents. Analyzing various factors like speed, acceleration, and following distance can reveal a driver's style. These styles can be categorized (e.g., aggressive, moderate, cautious) to better understand driving behaviour.

Autonomous vehicles need to anticipate the actions of surrounding cars. However, human drivers exhibit various driving styles, making it difficult to predict their movements. To navigate safely, autonomous vehicles must account for these unpredictable behaviours. Driving style profiling involves collecting data on a driver's actions and categorizing them based on specific parameters. This helps us understand their tendencies on the road. By recognizing driving styles, autonomous vehicles can better navigate mixed traffic environments with both human-driven and automated cars. Additionally, studying driving styles over time can reveal how factors like time of day influence behaviour. This knowledge can be used to improve overall road safety.

The challenge lies in using unlabeled data (data without predefined categories) to accurately predict driving style. This paper proposes an unsupervised learning approach to address this challenge for autonomous vehicles. Driving style, characterized by unique patterns of vehicle movement, is a key determinant of driver behaviour. This study focuses on the development of a driving style recognition system for intelligent vehicles using unsupervised clustering algorithms.

Previous research has highlighted the significance of driving style in traffic safety and autonomous vehicle development. For instance, [1] emphasizes the role of driving style in accident causation, suggesting that identifying and understanding these styles can contribute to accident prevention. Moreover, [2] underscores the challenges posed by diverse driving styles to autonomous vehicle trajectory prediction, emphasizing the need for robust models to account for these variations.

This paper proposes a model that uses machine learning to analyze vehicle trajectory data and classify drivers into three categories: aggressive, moderate, and traditional driving styles. This information can be used by autonomous vehicles to make safer decisions on the road. This approach is particularly advantageous as it eliminates the need for extensive data annotation, a time-consuming and resource-intensive process. The resulting clusters will represent different driving styles, enabling intelligent vehicles to better anticipate and respond to the behaviour of surrounding vehicles.

In the following sections, we will delve into the objective of our research which aims to develop a model for accurately categorizing driving styles surrounding autonomous vehicles, focusing on enhancing safety and efficiency in diverse traffic conditions, the next section, methodology involve data pre-processing, including cleaning and dimensionality reduction via Principal Component Analysis (PCA). The optimal number of driving style clusters was determined using the Elbow method and Silhouette analysis, followed by the application of the K-means clustering algorithm to categorize driving styles into aggressive, moderate, and traditional. Next section, result revealed distinct characteristics for each driving style. Finally in the discussion part we concluded the research that unsupervised clustering approach proved effective in classifying driving styles based on trajectory data, contributing to the understanding of driver behaviour and enhancing the capabilities of autonomous vehicles in mixed traffic environments. This research provides a novel framework for recognizing driving styles in autonomous vehicles, contributing to improved safety and decision-making in environments with mixed human and autonomous traffic.

2. Objectives

Accurately categorizing the driving behaviours of vehicles surrounding an autonomous system is essential for safe and efficient operation. Unsupervised clustering methods have emerged as a promising approach for this task due to their ability to discover hidden patterns within large datasets without requiring pre-labelled information. This section examines recent research exploring the application of unsupervised clustering techniques to classify driving styles of nearby vehicles.

Chen et al. (2019) employed a hierarchical clustering approach to classify driving styles of nearby vehicles. They utilized vehicle speed, acceleration, and deceleration data as input features. Their research identified four distinct driving style clusters: aggressive, normal, cautious, and hesitant [3].

Lee et al. (2018) employed a k-means clustering approach to classify driving styles of nearby vehicles. They utilized vehicle speed, acceleration, and lane-keeping data as input features. Their findings indicate that the k-means algorithm effectively differentiated between three primary driving styles: aggressive, normal, and conservative [4].

Wang et al. (2020) applied the DBSCAN clustering algorithm to classify driving styles of vehicles surrounding an autonomous vehicle. They utilized vehicle speed, acceleration, and lateral position relative to the autonomous vehicle as input features. Their findings indicated that DBSCAN effectively identified three primary driving style clusters: aggressive, normal, and cautious [5].

Furthermore, research has explored the combination of multiple clustering algorithms to enhance driving style classification accuracy. For instance, Zhang and colleagues (2021) implemented a hybrid approach integrating k-means and Gaussian Mixture Model (GMM) algorithms for categorizing driving styles. Their findings revealed that this combined method effectively distinguished four primary driving style clusters: aggressive, normal, cautious, and unpredictable [6].

Recent research has extensively employed unsupervised clustering algorithms for classifying the driving styles of vehicles surrounding autonomous vehicles. Commonly utilized methods include k-means, hierarchical clustering, DBSCAN, and hybrid combinations of these algorithms. These approaches have demonstrated effectiveness in distinguishing various driving styles, such as aggressive, normal, cautious, and hesitant.

The methodology presented in this paper introduces a novel framework that aims to improve the accuracy and efficiency of driving style classification, offering a significant advancement in addressing the complexities of this domain.

3. Methods

In this study, vehicle trajectory pairs are extracted from data collected on the I-80 and US-101 freeways. The NGSIM vehicle trajectory dataset, gathered from different regions and time periods, provides insights into both congested and moderate traffic conditions. Data pre-processing is a critical phase in this workflow. The pre-processing phase is done by identifying and addressing missing values, eliminating outliers to ensure the appropriate dataset to be used, and extracting relevant features [7]. The dataset's format is vital for analysis. Feature selection is performed using correlation plots, revealing that the data is randomly distributed across all comparison plots, indicating that all features are independent and influence the output label. The processed data is then input into the Google Colab platform using Python programming to achieve the desired outcomes. Machine learning algorithms are applied to the pre-processed dataset, and the results are analyzed. Figure 1 illustrates the method used for predicting driving styles.

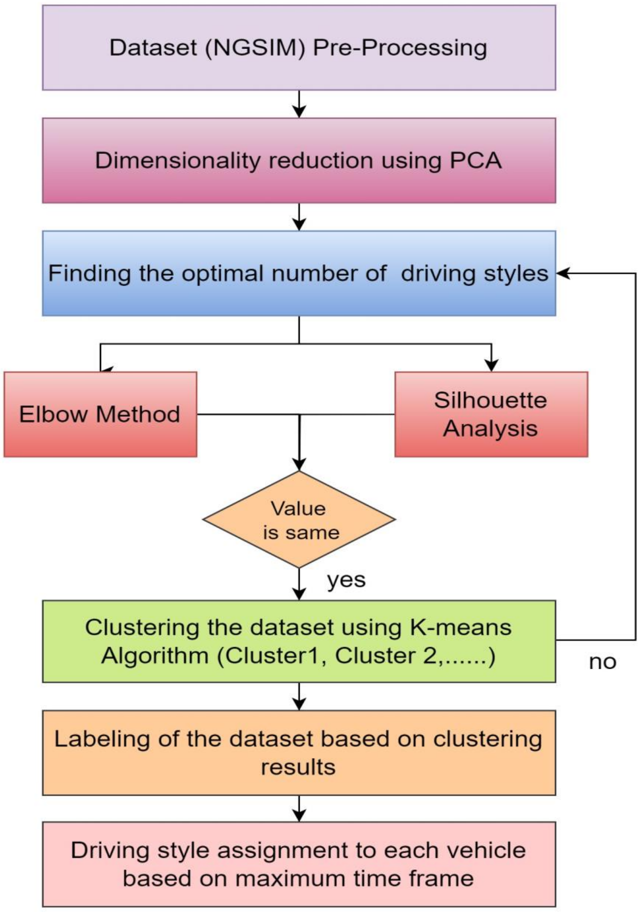

Figure 1: Method adopted for driving style prediction

3.1 Dataset Pre-processing:

The NGSIM dataset is perfect for analysing car-following behaviours since it is meticulously annotated with vehicle and frame identifiers [8]. The focus of this work is to aggregate leading-following pairs from different cars and environments. The existence of large inaccuracies in the raw NGSIM data is a major problem. For instance, a vehicle may create a collision path due to the natural trajectories in NGSIM, which is not compatible with how vehicles behave in the actual world [9]. A data preparation step is included to fix the original dataset in order to lessen this. The foundation is laid by the data cleaning procedure outlined in [10], with special care paid to interpolating implausible kinematic values using neighbouring data points [11], such as abrupt stops, unnaturally high acceleration rates, or sudden zero headway areas.

3.2 Dimensionality reduction using Principal Component Analysis (PCA)

Dimensionality reduction is applied to feature indicators to simplify the complex relationships within the data and to identify the key features that define driving styles. Principal Component Analysis (PCA), a widely used technique for reducing dimensions, creates new components by forming linear combinations based on the correlation matrix of the original variables. In PCA, a criterion is established to ensure that the selected components retain over 60% of the cumulative contribution rate, which reflects the amount of original characteristic information preserved [12]. After processing the trajectory data, PCA was performed, and the resulting cumulative contribution rate is shown in Figure 2.

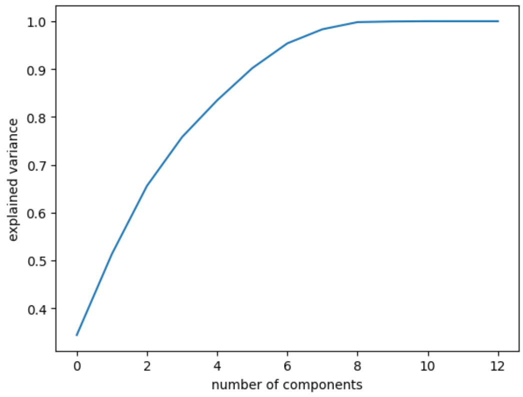

Figure 2: Principal component cumulative contribution rate (Components v/s variance)

By extracting the first two principal components, the cumulative contribution rate surpassed 60%, demonstrating that these two indices effectively capture the majority of information regarding merging vehicles [13], as illustrated in Figure 3.

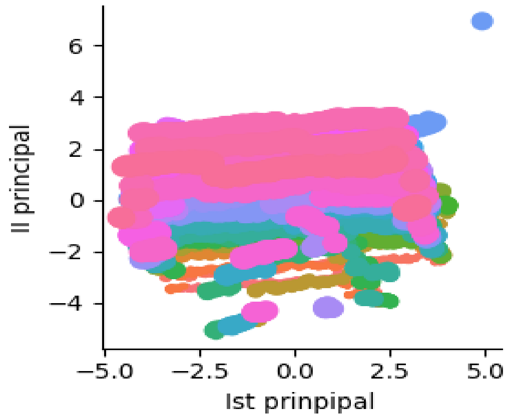

Figure 3: Dimensionality reduction of dataset in to 2 components using PCA

3.3 Finding optimal number of driving styles using Elbow method & Silhouette analysis

Driving styles are investigated through unsupervised classification using clustering analysis; the K-means clustering algorithm is a frequently utilised technique. This method is based on distance measurements, according to which an object's closeness to another indicates how similar they are. The basic idea is to allocate data points to different clusters iteratively. In this study, the distances between different samples and cluster centres were evaluated, and the K-means algorithm was employed to classify the combined vehicle data. Finding the K value, or ideal number of clusters, was the first step in the procedure. The "Elbow rule" and Silhouette value computations were used to evaluate the efficacy of various K values.

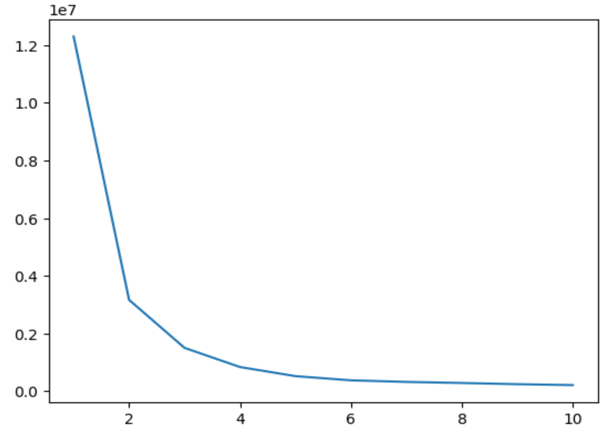

Figure 4: Finding clustering category number using Elbow method

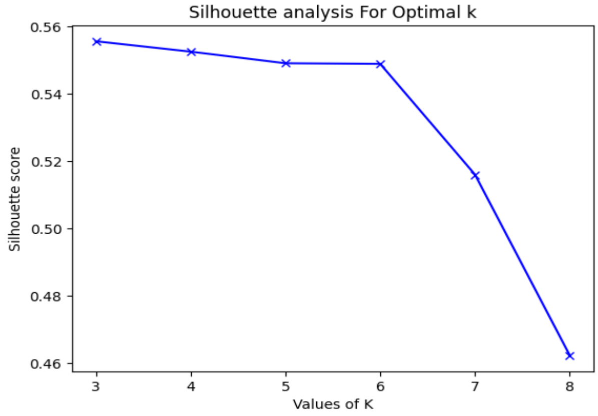

Figure 5: Finding clustering category number using silhouette analysis

Figure 4, which employs the Elbow Method, suggests that the optimal number of clusters k is around 3 or 4, as the sum of squared errors (SSE) shows a noticeable decrease before the curve begins to flatten. Figure 5, utilizing Silhouette Analysis, indicates that k=3 is the most suitable choice since it yields the highest silhouette score, reflecting well-defined and cohesive clusters. Together, these analyses suggest that k=3 is likely the optimal number of clusters.

3.4 Clustering the dataset by applying K-means Algorithm

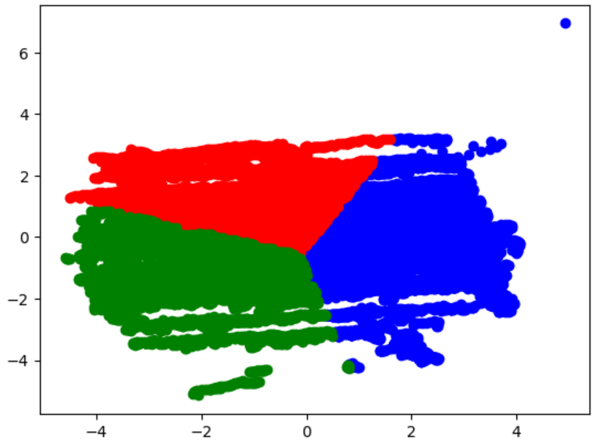

Figure 6: Dataset clustering by k-means algorithm

The clustering results, shown in Figure 6, highlight different driving styles based on characteristic index values. The classification identifies three distinct types: 0 represents aggressive driving, 1 indicates a moderate driving style, and 2 corresponds to a traditional driving style.

3.5 Labelling of the dataset & driving style assignment to the vehicle based on maximum time frame

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. The vehicle type is determined based on the label that appears most frequently for that vehicle. The label assigned to a vehicle corresponds to the label that has the highest number of occurrences or frames for that vehicle. Here, the time taken by vehicle in one frame is one tenth of a second. The procedure is detailed in Table 1.

4. Results

Three driving styles are distinguished for vehicles: aggressive, moderate, and traditional. In comparison to the other two categories, aggressive vehicles show faster speeds, more acceleration and jerk, shorter space headways, and shorter time headways (Table 2). This implies that drivers who are aggressive tend to drive faster and keep closer distances from the car in front of them. In addition, compared to the other two groups, aggressive vehicles had shorter time headway. When analysed based on metrics like velocity, acceleration, jerk, space headway, and time headway, moderate vehicles lie in between the aggressive and traditional classifications.

Table I: Driving style assignment to the vehicles based on maximum time frame

Vehicle ID	Total Time Frames	Time Frame in Label 0	Time Frame in Label 1	Time Frame in Label 2	Max. Time frame label
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 II: Comparison of aggressive, moderate and traditional driver

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 presents a comparison of aggressive, moderate, and traditional vehicles across various features, including velocity, acceleration, jerk, space headway, and time headway. The mean values for these features are used to compare the different vehicle categories. Aggressive vehicles have higher velocity, acceleration, and jerk, while their space and time headways are smaller compared to both moderate and traditional vehicles. In contrast, moderate vehicle values fall between those of aggressive and traditional vehicles. Traditional vehicles exhibit values that are the opposite of those for aggressive vehicles

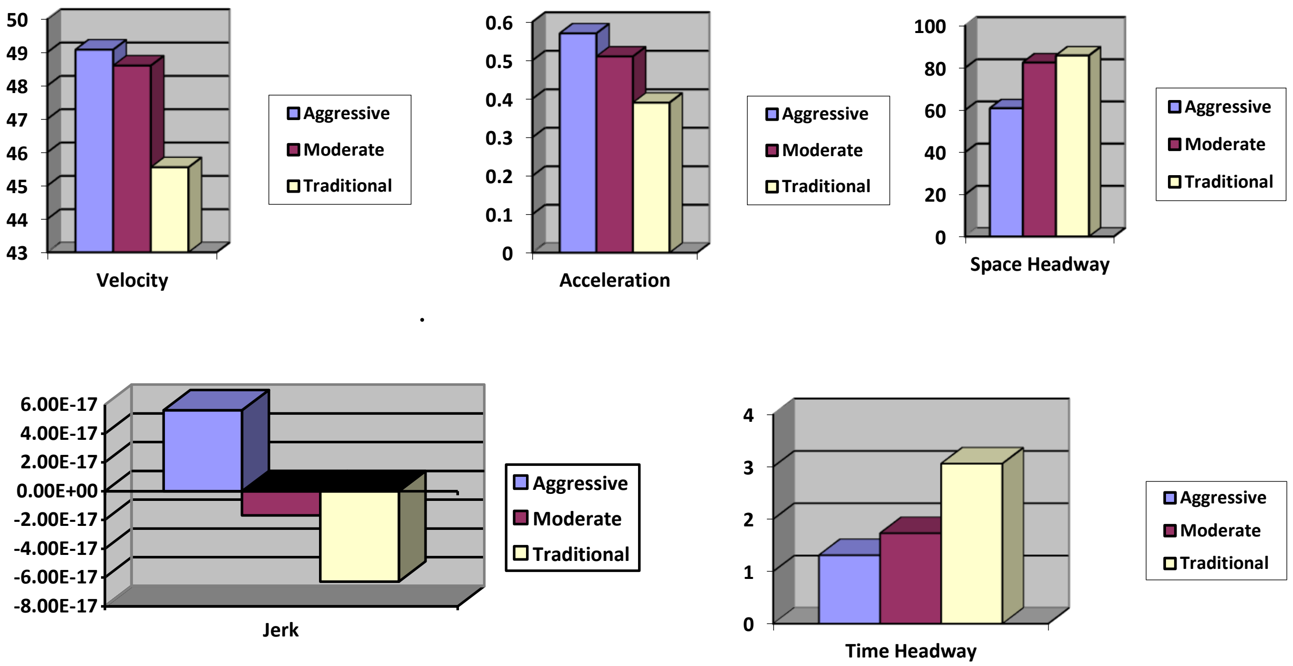

Figure.7: Driving Style Feature Comparison for Aggressive, Moderate, and Traditional Vehicles

Discussion

This study focused on categorizing vehicle driving styles by analyzing trajectory data. Principal Component Analysis (PCA) was employed to reduce the characteristic indexes to two principal components that encapsulate all relevant features. To determine the optimal number of categories, the "Elbow rule" and Silhouette methods were used, and K-means clustering was applied to group the driving styles. The resulting categories were aggressive, moderate, and traditional driving styles. The findings indicated that aggressive driving is associated with higher velocity, acceleration, and jerk, as well as shorter space and time headways. In future, the method for classifying driving styles based on unlabeled data can be expanded to include additional driving scenarios and new datasets.

Refrences

[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] 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.

[4] 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.

[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.