Identifying hard stop and momentary stop using vehicle trajectory dataset

1st Jay Raval, *Ahmedabad University*, Ahmedabad, India, jay.r1@ahduni.edu.in

2nd Meet Suthar, *Ahmedabad University*, Ahmedabad, India, meet.s4@ahduni.edu.in

3rd Karan Prajapati, *Ahmedabad University*, Ahmedabad, India, karan.p1@ahduni.edu.in

4th Jinil Savaj, *Ahmedabad University*, Ahmedabad, India, jinil.s@ahduni.edu.in

5th Vishesh Bhatia, *Ahmedabad University*, Ahmedabad, India, vishesh.b@ahduni.edu.in

***Abstract*—This document presents a machine learning project for classifying vehicle behavior at signalized intersections using aerial-view trajectory data. Vehicles are classified as hard stops, momentary stops, or moving based on position, velocity, and acceleration. A Random Forest classifier, trained with preprocessing and feature engineering, is evaluated using accuracy, precision, recall and visualization through time-series plots to enhance traffic monitoring and signal optimization.**

***Index Terms*—Machine learning, vehicle behavior classification, Trajectory Analysis, Traffic Monitoring, Random forest, Clustering**

I. Introduction

Classifying vehicle behavior at signalized intersections is essential for traffic analysis, safety, and autonomous vehicle research. This project classifies vehicles into hard stops, momentary stops, and moving vehicles using trajectory data from aerial-view videos. The dataset captures position, velocity, and acceleration. Machine learning analyzes movement patterns, while visualizations validate results, enhancing traffic monitoring, signal optimization, and urban mobility.

II. Methodology

A. Dataset Description

The dataset, extracted from aerial-view videos of signalized intersections, tracks vehicle trajectories with attributes such as frame number, vehicle ID, and bounding box coordinates (left, top, width, height). Table I illustrates sample data. Preprocessing steps ensured data consistency by handling missing values and standardizing measurements for reliable analysis.

TABLE I
Sample Vehicle Tracking Data

frameNo	trackID	left	top	w	h
1	155	2019	417	63	30
1	154	2018	376	61	30
1	153	949	1267	56	51
1	152	1426	1596	71	54
1	151	2466	1534	38	40

B. Feature Engineering

Feature engineering enhances vehicle behavior classification by extracting meaningful insights from raw trajectory data.

Center X	Center Y	Overall CLS	Distance	Velocity	Acceleration	Speed	Avg Velocity 50	Avg Velocity 100
250	1027.5	Car	1000.04	2.45	NaN	25.26	3.0048	3.1587
256.5	1026.5	Car	998.58	-1.46	-3.92	23.68	3.0176	3.632
265.5	1027.5	Car	1002.29	3.72	5.18	32.6	3.1547	3.562
......								
......								
273	1028	Car	1003.76	1.47	-2.25	27.06	3.1341	3.0058
280	1027.5	Car	1005.29	1.53	0.06	25.26	3.1696	3.8912
304	1028.5	Car	1011.22	0.52	-2.32	27.24	3.1578	3.7541

Fig. 1. Corrected Dataset

1) Key Features:

- **Velocity:** Measured using the displacement of vehicle center positions across frames, providing insights into movement speed.
- **Acceleration:** Derived from velocity changes over time, helping to identify sudden stops and gradual slowdowns.
- **Stopping Behavior:** Vehicles are classified into three categories:
  - **Hard Stop:** Permanently stationary vehicles.
  - **Momentary Stop:** Brief pauses before resuming motion.
  - **Moving Vehicle:** Continuous movement without stops.

2) Refined Dataset: The dataset was corrected to ensure accurate tracking of vehicle attributes (Fig. 1).
It includes Center X/Y for position, Overall CLS for object classification, and Distance to track movement. Velocity and Acceleration capture speed and sudden changes, helping identify stops. Speed provides a filtered value for stability, while Avg Velocity 50/100 smooths speed trends over multiple frames. These refinements improve motion analysis, leading to more reliable classification.

C. Machine Learning Model

1) Model Training: The dataset was split into 80% training and 20% testing to ensure fair evaluation. A Random Forest Classifier was trained using extracted features, with hyperparameter tuning applied to optimize performance. Key

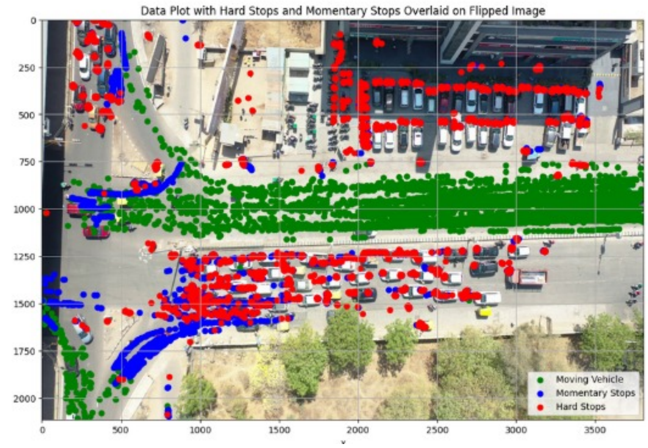

Fig. 2. Traffic Analysis Overhead View: Visualization of Moving Vehicles (green), Momentary Stops (blue), and Hard Stops (red) overlaid on a flipped aerial image for traffic behavior assessment.

parameters adjusted include the number of trees (to balance bias and variance), maximum depth (to prevent overfitting), and minimum samples per split (ensuring meaningful decision-making).

2) Advantages of Random Forest: Random Forest was chosen for its ensemble learning approach, making it resistant to overfitting while effectively handling complex, non-linear feature relationships. Additionally, it provides interpretability by highlighting feature importance, helping understand how different parameters contribute to classification.

D. Evaluation Metrics

The model's effectiveness was assessed using standard classification metrics:

- **Accuracy:** Measures overall correctness.
- **Precision:** Ensures predicted stops are correct, reducing false alarms.
- **Recall:** Measures the ability to identify all real stop cases.
- **F1-Score:** Balances precision and recall, useful for imbalanced data.

E. Visualization Techniques

Various visualization techniques were employed to validate results and interpret predictions:

- **Scatter Plots:** Used to analyze vehicle positions and movement patterns.
- **Coordinate-Based Mapping:** Displays classified stops on an intersection layout for spatial analysis.
- **Image Overlays:** Superimposes classification results on real-world images, confirming prediction accuracy.

These techniques helped refine the model's performance and provided insights into traffic patterns for real-world applications.

III. Result

The trained Random Forest model achieved an overall accuracy of 85%, effectively classifying vehicle behavior at signalized intersections. The precision, recall, and F1-score metrics for each category are presented in Table II.

TABLE II
MODEL PERFORMANCE METRICS

Class	Precision	Recall	F1-Score	Support
Moving (0)	0.89	0.92	0.91	87,718
Momentary Stop (1)	0.84	0.87	0.85	14,991
Hard Stop (2)	0.52	0.39	0.44	13,361
Accuracy	-	-	0.85	116,070
Macro Average	0.75	0.73	0.73	116,070
Weighted Average	0.84	0.85	0.85	116,070

Visualization techniques, including scatter plots and coordinate-based mapping, validated the model's predictions. Image overlays on real-world intersections further confirmed the accuracy of the classifications, revealing clear spatial distribution patterns for stopping behavior.

IV. Discussions

The model demonstrated strong performance in classifying moving vehicles, achieving an F1-score of 0.91 with high precision (0.89) and recall (0.92). Momentary stops were also well detected, with an F1-score of 0.85. However, classifying hard stops proved more challenging, with a lower F1-score of 0.44, precision of 0.52, and recall of 0.39. This suggests that hard stops share overlapping characteristics with momentary stops, making them harder to distinguish accurately.

Despite strong performance, misclassifications of hard stops suggest the need for better feature engineering or data labeling. Future improvements could incorporate traffic signals and road conditions to enhance accuracy.

References

[1] V. Patel, M. Maleki, Mehdi Kargar, J. Chen, and H. Maoh, "A cluster-driven classification approach to truck stop location identification using passive GPS data," Journal of Geographical Systems, vol. 24, no. 4, pp. 1–21, Jun. 2022, doi: https://doi.org/10.1007/s10109-022-00380-y

[2] Negin Masnabadi, Farhad Hosseinali, and Zahra Bahramian, "Developing a spatial and temporal density-based clustering algorithm to extract stop locations from the user's trajectory," Journal of Geospatial Information Technology, vol. 9, no. 2, pp. 105–128, Oct. 2021, doi: https://doi.org/10.52547/jgit.9.2.105.

[3] Y. Hu, Y. Li, H. Huang, J. Lee, C. Yuan, and G. Zou, "A high-resolution trajectory data driven method for real-time evaluation of traffic safety," Accident Analysis & Prevention, vol. 165, p. 106503, Feb. 2022, doi: https://doi.org/10.1016/j.aap.2021.106503.

[4] G. Y. Oukawa, P. Krecl, and A. C. Targino, "Fine-scale modeling of the urban heat island: A comparison of multiple linear regression and random forest approaches," Science of The Total Environment, vol. 815, p. 152836, Apr. 2022, doi: https://doi.org/10.1016/j.scitotenv.2021.152836.