Ahmedabad University

CSE523 : Machine Learning | The Overfitters

Identifying hard stop & momentary stop using vehicle trajectory dataset

Vishesh Bhatia
AU2240027

Jay Raval
AU2240151

Jinil Savaj
AU2240159

Karan Prajapati
AU2240161

Meet Suthar
AU2240198

Problem Statement

- Given a dataset containing extracted vehicle trajectories at a signalized intersection to analyze different stopping behaviors.
- At intersections, we assume drivers follow traffic signals, leading to vehicles stopping and restarting at red lights, which we classify as a momentary stop.
- Some vehicles remain stationary for an extended period, such as parked vehicles, representing a hard stop.
- There are vehicles that continue moving without stopping, categorized as moving vehicles.
- Our objective is to develop a machine learning model to accurately distinguish between these three conditions using trajectory data.

Literature Survey

Author	Published Year	Publication Title	Key Findings
V. Patel, M. Maleki, Mehdi Kargar, J. Chen, and H. Maoh	June. 2022	A cluster-driven classification approach to truck stop location identification using passive GPS data	DBSCAN clustering for stop identification, feature-based classification, Random Forest classifier achieving 97% accuracy for primary stops, 83% overall accuracy, and 92% prediction accuracy for primary stops.
Negin Masnabadi, Farhad Hosseinali, and Zahra Bahramian	October. 2021	Developing a spatial and temporal density-based clustering algorithm to extract stop locations from the user's trajectory	Enhanced DBSCAN with spatial-temporal indices, solved round trip issue, improved cluster detection, and achieved 94.06% accuracy in identifying holidays.

Dataset Discussion

Initial Dataset Without Correction

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
1	150	337	0	63	48
1	149	1462	846	82	54
1	148	1855	128	62	36
1	147	2023	1035	38	32

- **frameNo:** Represents the frame number in the video sequence. Each frame corresponds to a single image captured at a specific timestamp.
- **trackID:** A unique identifier assigned to each detected vehicle. Useful for tracking the trajectory of each vehicle over time.
- **left:** The x-coordinate (in pixels) of the top-left corner of the vehicle's bounding box.
- **top:** The y-coordinate (in pixels) of the top-left corner of the vehicle's bounding box.
- **w (width):** The width of the bounding box (in pixels).
- **h (height):** The height of the bounding box (in pixels).

Dataset Discussion

Final Dataset With Correction

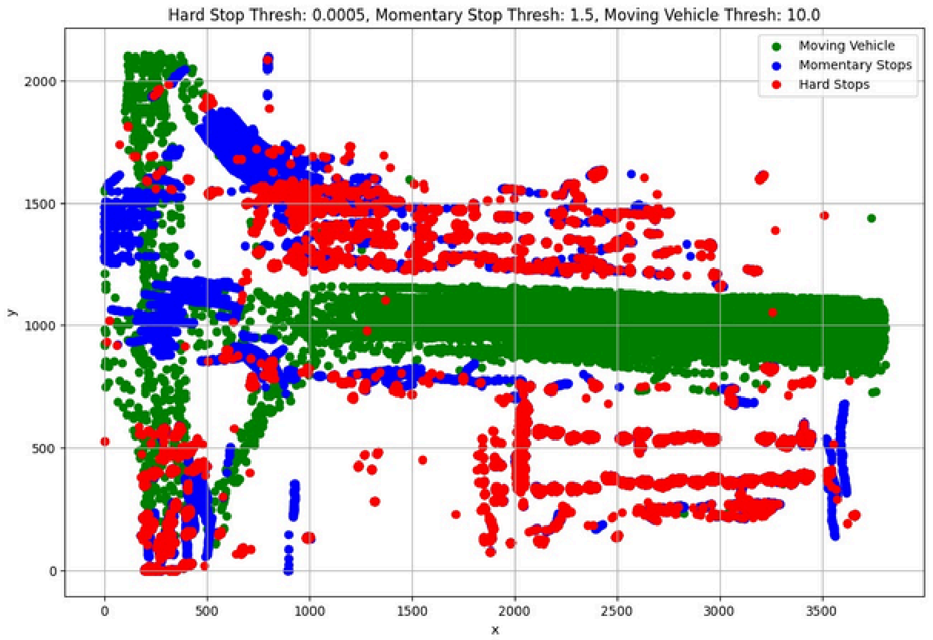
Center X	Center Y	Overall CLS	Distance	Velocity	Acceleration	Speed	Avg Velocity 50	Avg Velocity 100
250	1027.5	Car	1000.04	2.45	NaN	25.26	3.0048	3.1587
256.5	1026.5	Car	998.58	-1.46	-3.92	23.68	3.0176	3.632
265.5	1027.5	Car	1002.29	3.72	5.18	32.6	3.1547	3.562
……								
……								
273	1028	Car	1003.76	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

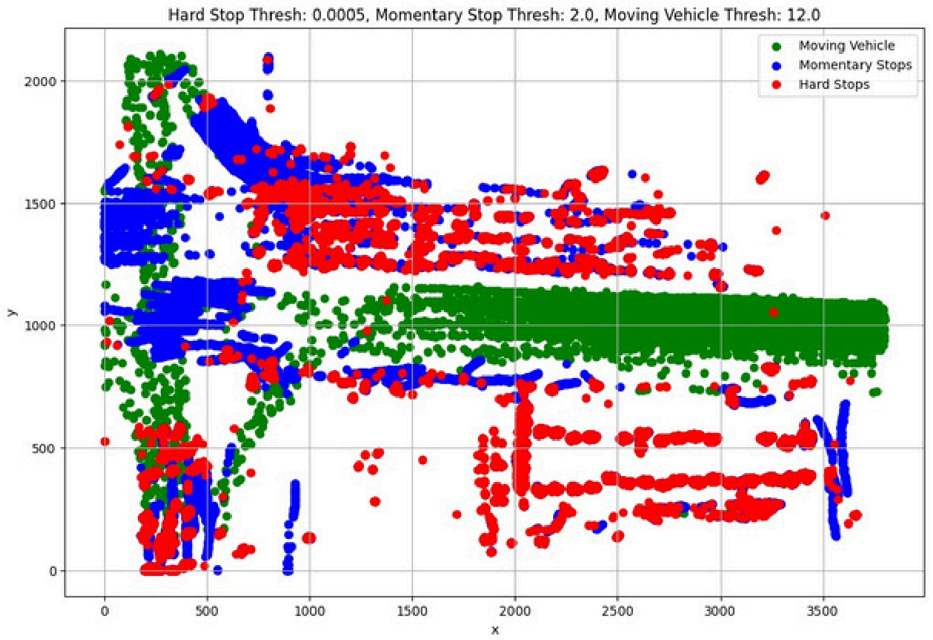
- **Center X / Center Y:** Represent the position coordinates of the vehicle in the frame.
- **Overall CLS:** Identifies the type of object.
- **Distance:** Represents the distance covered by the vehicle.
- **Velocity:** Measures the vehicle's current speed at a given moment.
- **Acceleration:** Tracks changes in velocity, which helps in identifying sudden stops or quick acceleration.
- **Speed:** Similar to velocity but represent average speed or filtered values.
- **Avg Velocity 50 / Avg Velocity 100:** Represent averaged velocity readings over the last 50 and 100 samples, respectively, providing smoother data trends.

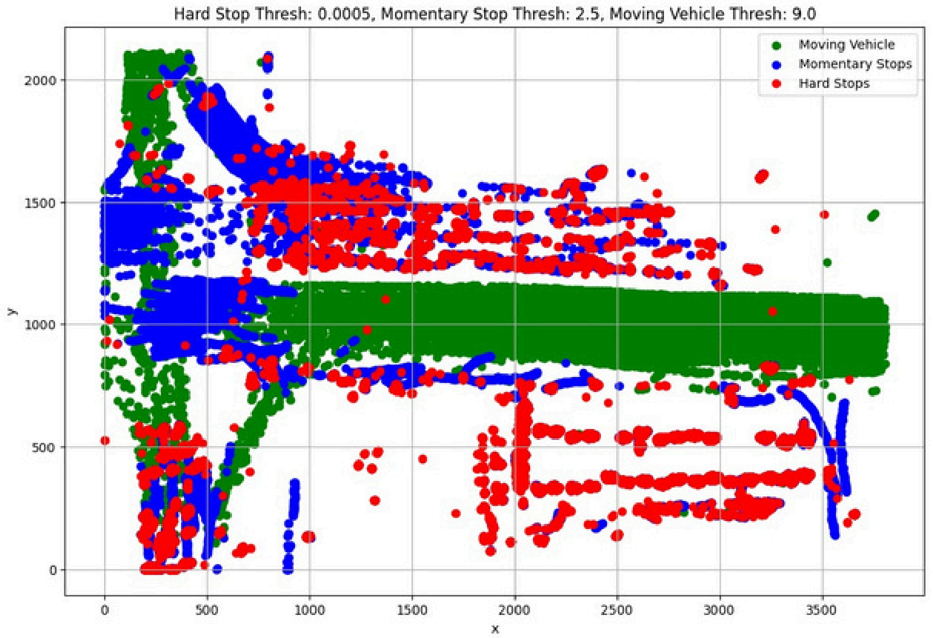
Approach

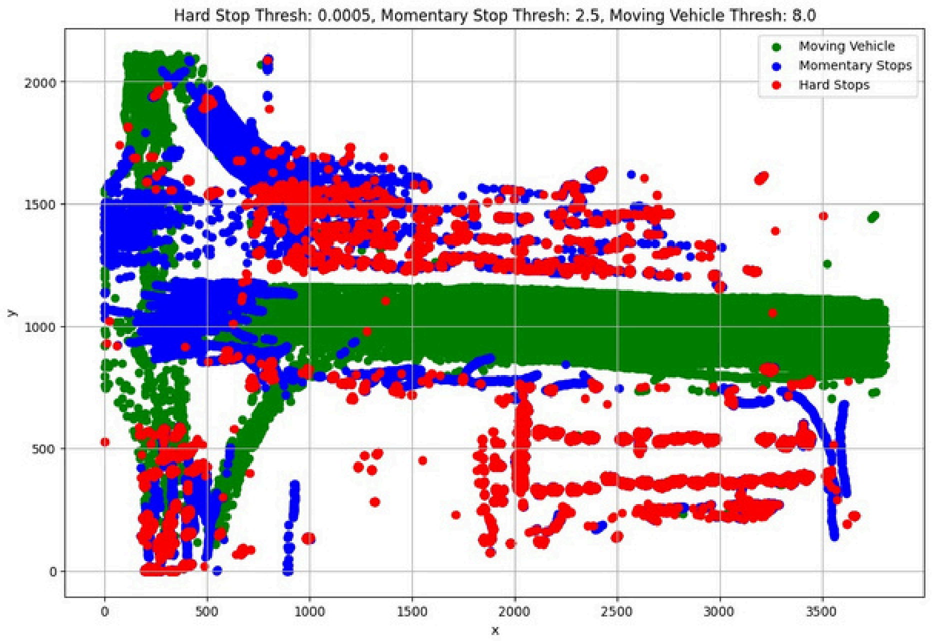
We've plotted the graph for several thresholds values and after comparing these plots with prior beliefs having seen the video, we found these thresholds appropriate.

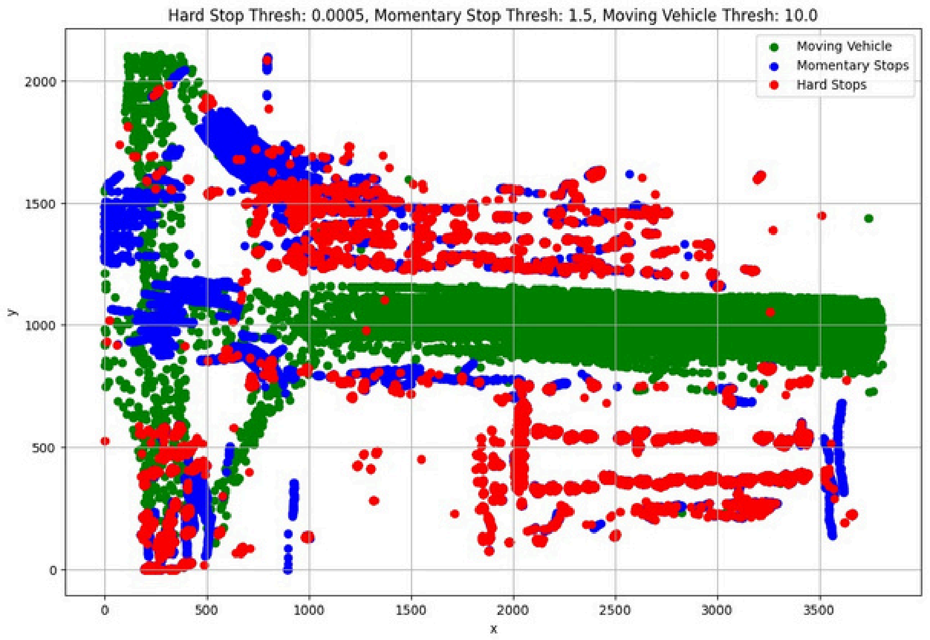
```
hard_stop_thresholds = [0.0005, 0.0007, 0.0009, 0.001, 0.0011, 0.0013, 0.0015]
momentary_stop_thresholds = [1.5, 1.7, 1.9, 2.0, 2.1, 2.3, 2.5]
moving_vehicle_thresholds = [8.0, 9.0, 10.0, 11.0, 12.0]
```

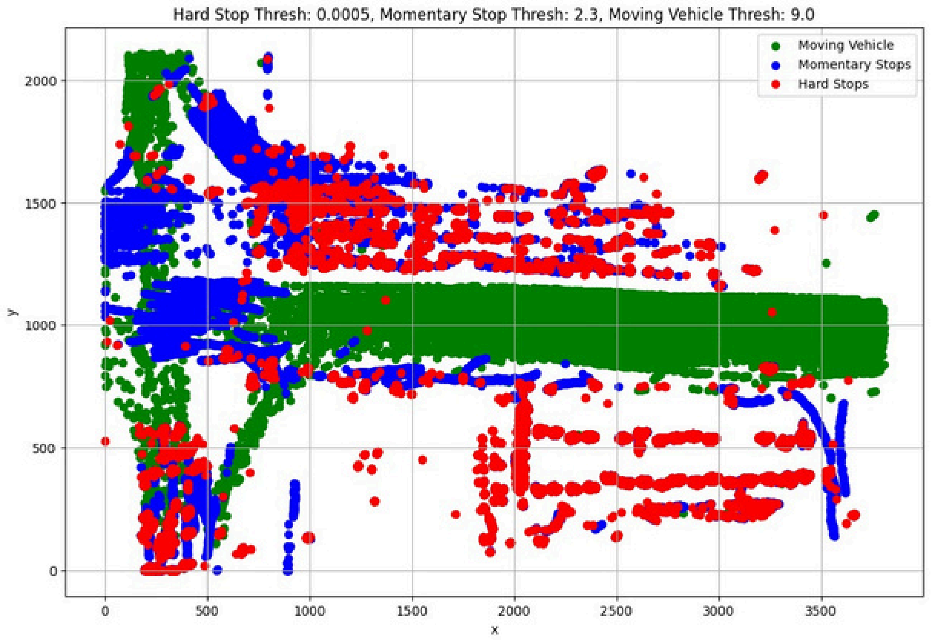








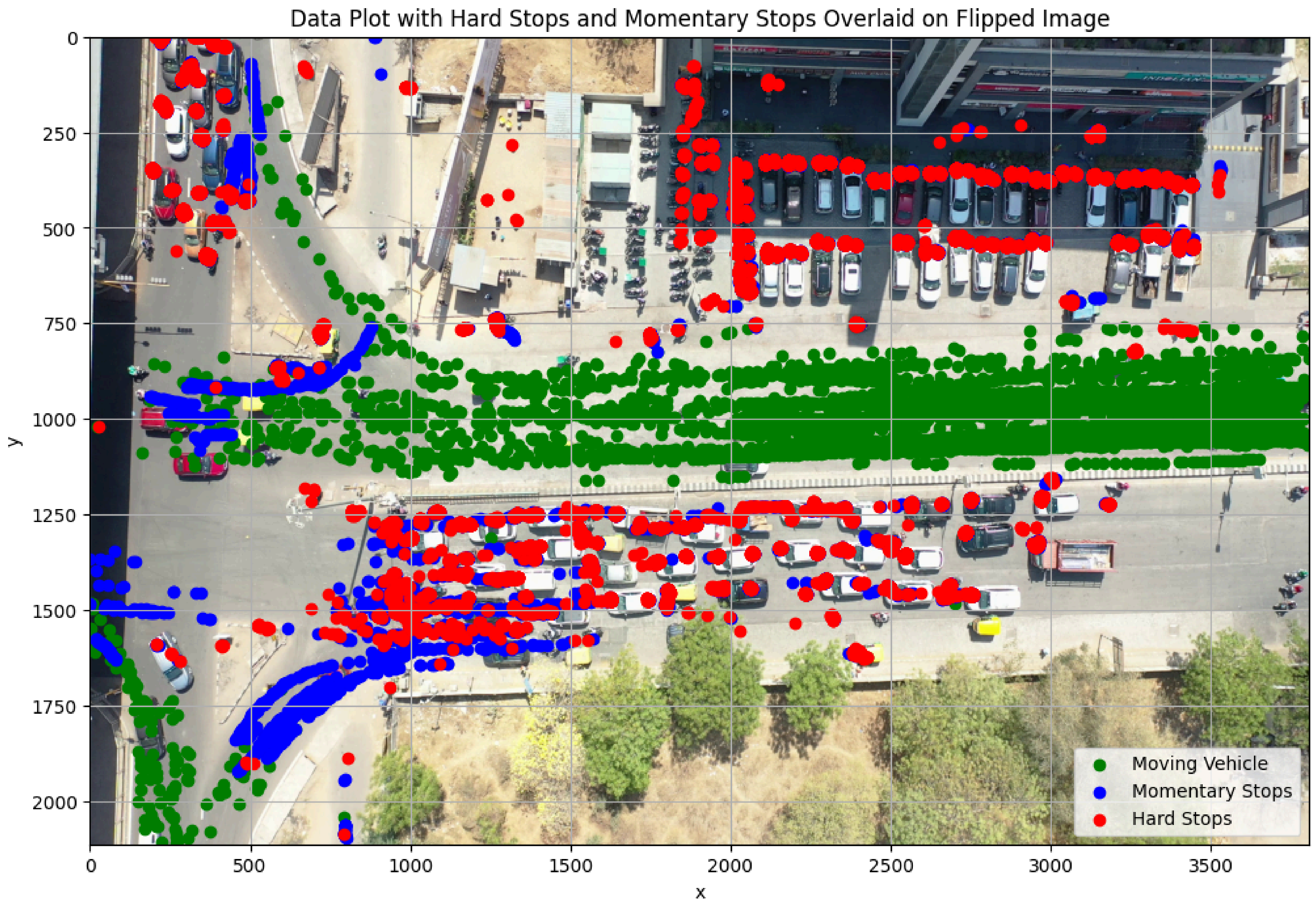




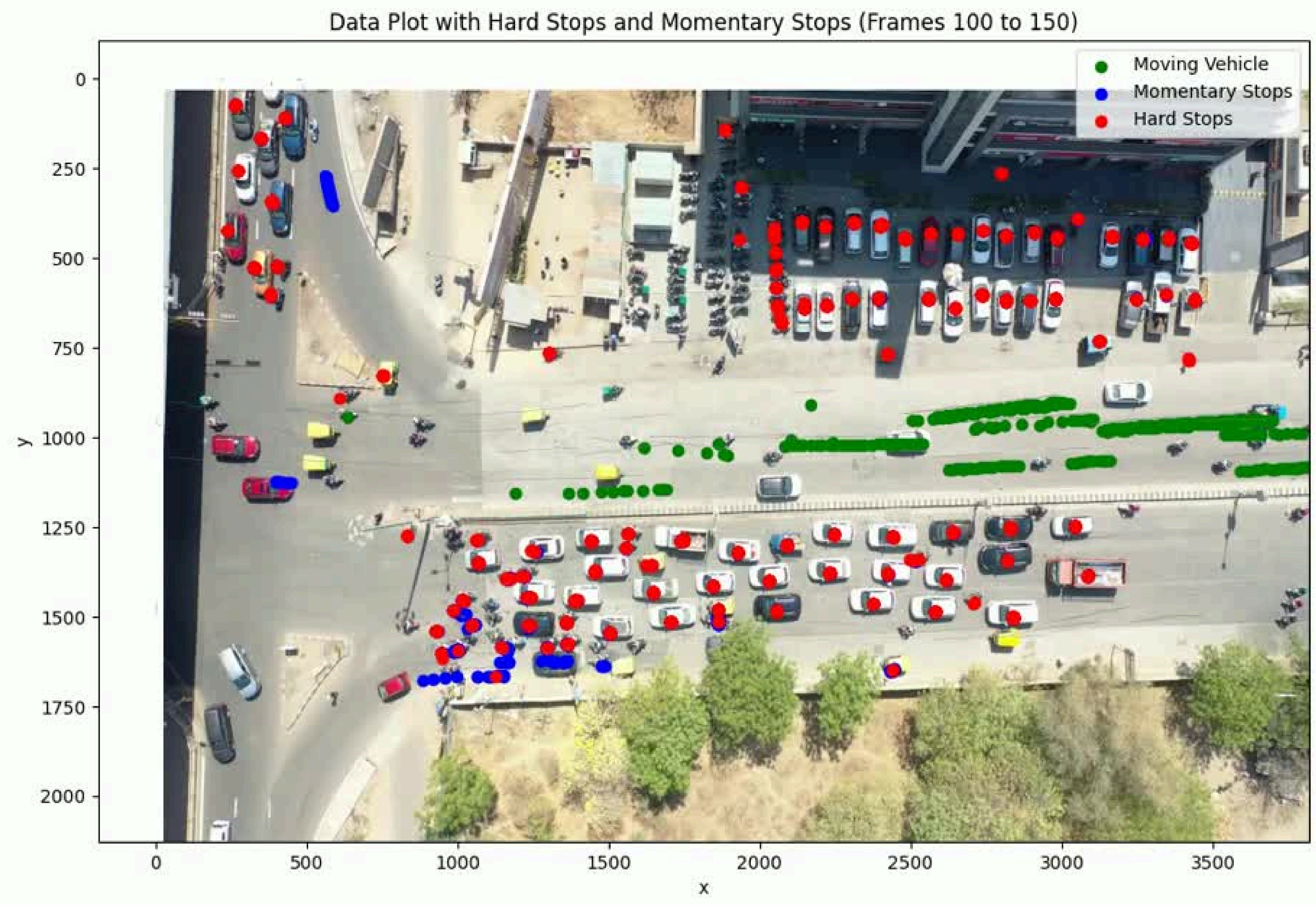
Approach

- The Random Forest model was used for classification, leveraging multiple decision trees to improve accuracy and reduce overfitting.
- It was trained on **distance**, **velocity**, and **acceleration**, with missing values handled using SimpleImputer.
- The dataset was split 80%-20% for training and testing.
- Performance was evaluated using accuracy score, classification report, and confusion matrix, and results were visualized to analyze vehicle movement and stop patterns.

```
hard_stop_threshold = 0.001
momentary_stop_threshold = 2.0
moving_vehicle_threshold = 10.0
```



Results




```
Accuracy Score: 0.8529680365296803
Classification Report:
              precision    recall  f1-score   support

           0       0.89      0.92      0.91     87718
           1       0.84      0.87      0.85     14991
           2       0.52      0.39      0.44     13361

    accuracy                           0.85    116070
   macro avg       0.75      0.73      0.73    116070
weighted avg       0.84      0.85      0.85    116070
```

Future Work

- **Exploring Alternative Methods for Data Labeling:** Investigate alternative methods to assign labels, anomaly detection, or supervised learning to improve labeling accuracy without relying on fixed thresholds.

- **Fine-Tuning Threshold Values for Optimal Labeling:** Optimize threshold values using data-driven techniques like grid search or Bayesian optimization for more precise classification.

- **Comparing Model Performance with DBSCAN:** Evaluate DBSCAN’s ability to classify vehicle stops without predefined thresholds and compare its accuracy with the current Random Forest model.

References

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