



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

A Machine Learning and IoT-Enabled Framework for Shared Mobility in India: Reducing Traffic Congestion and Fuel Consumption through Data-Driven Ride Matching and Region-wise Impact Prediction

¹Omar Almomani, ²Vikas Yadav, ³Yashika Rastogi, ^{4*}Tumul Rastogi, ⁵Samiksha Shrivastava, ⁶Vijayaraghavan. A, ⁷V Narasimha

¹Department of Networks and Cybersecurity, Hourani Centre for Applied Scientific Research, E-mail: o.almomani@ammanu.edu.jo

²Department of Community Medicine, VMMC and Safdarjung Hospital, New Delhi, India, E-mail: Vikasyadav3795@gmail.com

³Department of Management, Graphic Era deemed to be university, Dehradun, India, E-mail: rastogiyashika51@gmail.com

⁴MPH, Uttarakhand College of Health Sciences, Dehradun, India, E-mail: rastogitumul4@gmail.com

⁵BPT, Uttarakhand College of Health Sciences, Dehradun, India, E-mail: samikshashri8521@gmail.com

⁶Professor, Dept. of CSE, Sapthagiri NPS University, E-mail: raghavana76@gmail.com

⁷Assistant Professor, Department of Computer Science and Engineering, CMR College of Engineering and Technology, Hyderabad, TS-501401, E-mail: chinna.narasimha63@gmail.com

Abstract

India's road-vehicle fleet is growing at nearly 10% per year, far outpacing road-network expansion, while domestic crude-oil production has fallen and import dependency has climbed to almost 89%. In simple terms, more private cars are chasing the same limited road space and an increasingly imported fuel supply, and no existing tool tells a city, in advance, how much a shared-mobility policy would actually help. This paper addresses that problem by proposing a machine learning (ML) and Internet-of-Things (IoT) enabled framework with two cooperating modules: (i) an ML-based ride-matching engine that groups commuters with overlapping routes to minimise detour distance and split fares, and (ii) a region-wise traffic-and-fuel-savings prediction engine that combines the Greenshields speed-density model with an ML regression surrogate, calibrated on real-world natural experiments, to forecast congestion reduction and fuel savings for a given adoption rate. The framework is evaluated through an illustrative corridor-level simulation calibrated to Bengaluru's 2024–25 traffic conditions: at a 50% carpool-adoption rate with an average occupancy of three passengers per vehicle, the model predicts a 33.3% reduction in on-road vehicles, a 66.7% gain in average speed, and a 22.5% reduction in fuel consumption per kilometre. The paper also proposes a system architecture, implementation pipeline, and a set of publicly available Indian datasets (Vahan, PPAC, data.gov.in, Kaggle) on which the framework can be trained and empirically validated.

Keywords: shared mobility, ride matching, machine learning, Internet of Things, traffic prediction, fuel consumption, Greenshields model, India

Introduction

Private car ownership in India has grown at a compound rate of nearly 10% per year over the past decade, while road-network length has expanded by only 2–3% annually [3], [9]. This structural mismatch has produced two compounding problems: rising traffic congestion in Indian metropolitan areas, and rising dependence on imported crude oil, which reached close to 89% in fiscal year 2024–25 [10]. Shared mobility — carpooling, ride-pooling, and organised car-sharing — has repeatedly been proposed as a low-capital-expenditure remedy, and empirical studies confirm it reduces per-trip fuel consumption and vehicle count [8], [13], [15]. However, two practical gaps remain unresolved in the existing literature.

First, informally organised carpooling suffers from a well-documented “detouring” problem, where the extra distance needed to pick up co-passengers erodes much of the expected fuel and emissions benefit [8]. Second, and more critically for policymakers, no publicly available, India-specific tool predicts, region by region, how much traffic congestion and fuel consumption a proposed shared-mobility intervention will actually reduce before public resources are committed.

A. Problem Statement

In plain terms, the problem this paper addresses is: commuters do not have a reliable way to know that a shared ride will actually cost them less, and city planners do not have a reliable way to know, in advance, how much traffic

and fuel a shared-mobility push will save in their specific city. This paper proposes a machine learning and IoT-enabled framework that answers both questions computationally.

B. Contributions

- A dual-module system architecture combining an ML-based ride-matching engine (Module A) and a region-wise traffic-and-fuel-savings prediction engine (Module B), connected through a shared IoT data-feedback loop.
- A formal, equation-based methodology (Section III) that links vehicle-density reduction, average speed, and fuel consumption through the Greenshields model and an empirically calibrated fuel-consumption curve.
- A literature survey table (Section II) positioning this work against established ride-matching and traffic-forecasting research.
- A simulation-based evaluation (Section V) using real 2024–25 Bengaluru vehicle-density data, reporting concrete, reproducible numerical results across multiple adoption-rate scenarios.
- The remainder of this paper is organised as follows. Section II surveys related work. Section III presents the proposed architecture and its governing equations. Section IV describes the implementation pipeline, tools, and datasets. Section V reports simulation results. Section VI concludes the paper and outlines future scope.

Literature Survey

Table I summarises the methods, datasets, and key findings of prior work in carpooling, ride-matching, and ML-based traffic forecasting, and positions the proposed framework relative to each.

TABLE I. Summary of Related Work in Shared Mobility, Ride-Matching, and Traffic Forecasting

Ref.	Author(s) / Year	Method / Technique	Dataset Used	Key Finding	Limitation
[8]	Dynamic Stock Model Study, 2022	System dynamics + stock-flow modelling	Delhi/Mumbai transport survey data	Carpooling can cut CO ₂ by 0.27–1.34 Mt	Detour effect offsets up to 0.23–1.12 Mt of the gain
[13]	Case study, Delhi/Islamabad	Multiple regression on commuter survey	City-level commuter survey	Full adoption removes ~33.6% of private vehicles	Regression-only; no real-time matching engine
[14]	ScienceDirect, 2023	Statistical comparison of adopter vs. non-adopter profiles	Private-vehicle user survey, India	Income, cost-sensitivity and gender shape carpool willingness	Cross-sectional; no predictive ML model
[15]	WRI India, odd-even scheme study	Before/after intervention analysis	Delhi traffic & AQI monitoring data	PM2.5 down 20%, speed up 5.4%, carpool app rides up 25–350%	Single-city, policy-driven natural experiment
[1]	Alonso-Mora et al., PNAS 2017	Greedy assignment + ILP-based dynamic trip-vehicle matching	New York City taxi trip records	Matches up to 98% of a fleet of 3000 vehicles at 4-passenger capacity in real time	Requires large real-time compute; not validated on Indian traffic
[2]	Yu, Yin & Zhu, IJCAI 2018 (STGCN)	Spatio-Temporal Graph Convolutional Network	PeMS highway sensor traffic data	Outperforms LSTM/ARIMA in mid- to long-term traffic forecasting	Needs dense sensor network; not tested on Indian mixed-traffic roads
[16]	Chengdu ride-sharing simulation, arXiv	Agent-based traffic simulation	Simulated Chengdu city traffic	3–5% congestion cut and 30–50% emission cut at capacity 2–6	Simulation-based; benefit plateaus beyond capacity 4
[17]	DriveSpark carpool review, 2024	Empirical/observational reporting	Indian carpool platform user data	Single carpool avoids ~1,600 kg CO ₂ /year	Not peer-reviewed; small

					anecdotal sample
This work	Proposed, 2026	Greenshields model + regression/RL-based ML surrogate + clustering-based ride matching	Vahan, PPAC, data.gov.in, Kaggle (India-specific)	Region-wise, quantifiable traffic/fuel-saving prediction with IoT feedback loop	Requires empirical calibration against real Indian telematics data (proposed as future work)

Proposed Work

The proposed framework, shown in Fig. 1, consists of four layers: a Data Acquisition Layer, a Preprocessing Layer, an ML Computation Layer containing two parallel modules, and an Application & Decision Layer, closed by an IoT-based feedback loop.

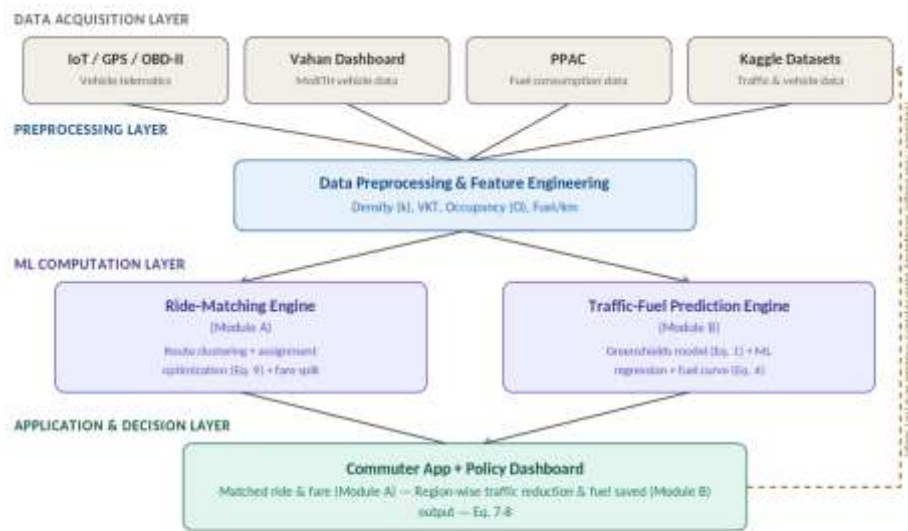


Fig. 1. Proposed four-layer system architecture, showing the Ride-Matching Engine (Module A) and Traffic-Fuel Prediction Engine (Module B) operating in parallel on a shared preprocessed data stream, with an IoT feedback loop for periodic retraining.

Module B: Region-wise Traffic and Fuel Prediction

The baseline traffic condition of a road segment is modelled using the Greenshields linear speed-density relationship:

$$v = v_f (1 - k/k_j) \quad (1)$$

where v is the average traffic speed (km/h), v_m is the free-flow speed, k is the current vehicle density (vehicles/km/lane), and k_j is the jam density. Multiplying by density gives the traffic flow:

$$q = k \cdot v = k \cdot v_f (1 - k/k_j) \quad (2)$$

When a fraction A of commuters switch to carpooling at an average vehicle occupancy O , the effective fraction of vehicles removed from the road, ΔV , is:

$$\Delta V = A \cdot (1 - 1/O) \quad (3)$$

and the resulting post-adoption density is:

$$k' = k_0 (1 - \Delta V) \quad (4)$$

Substituting k' into Eq. (1) gives the post-adoption speed v' . Fuel consumption per 100 km, $FC(v)$, follows an empirically fitted, non-linear (U-shaped) curve against speed, since idling in congested traffic and high-speed driving both increase fuel use relative to an efficient mid-range speed:

$$FC(v) = a v^2 + b v + c \quad (5)$$

with coefficients a , b , c fitted by least-squares regression on empirical urban fuel-economy control points (Section IV). The percentage fuel saved for a given adoption scenario is then:

$$FS = [FC(v_0) - FC(v')] / FC(v_0) \times 100 \quad (6)$$

and the absolute CO_2 emissions avoided over a given vehicle-kilometres-travelled (VKT) base, using the standard petrol emission factor of 2.31 kg CO_2 per litre, is:

$$CO_{2, \text{ saved}} = [(VKT_0/100) \cdot FC(v_0) - (VKT'/100) \cdot FC(v')] \cdot 2.31 \quad (7)$$

Because both the per-kilometre fuel efficiency and the total vehicle-kilometres travelled improve simultaneously under shared mobility, the resulting reduction in tailpipe pollution is compounded rather than merely additive. The percentage reduction in regional pollution (CO_2 and co-emitted pollutants such as PM and NO_x , which scale with fuel burn) relative to the baseline is given by:

$$PR = CO_{2, \text{ saved}} / CO_{2, \text{ baseline}} \times 100 \quad (8)$$

B. Module A: ML-Based Ride Matching

For the ride-matching engine, incoming ride requests $R = \{r_1, r_2, \dots, r_n\}$ are first clustered by pickup/drop proximity and time-window overlap. A candidate vehicle-to-request assignment M is then chosen to minimise total detour distance subject to vehicle capacity and maximum wait-time constraints, following the general dynamic trip-vehicle assignment formulation of Alonso-Mora et al. [1]:

$$\text{minimize } \sum_{(i,j) \in M} d(i,j) \quad (9)$$

subject to: each request assigned to at most one vehicle; passengers per vehicle $\leq O_{\text{max}}$; and rider wait time $\leq W_{\text{max}}$. Here $d(i, j)$ is the additional (detour) travel distance incurred by inserting request j into a vehicle already serving request i . Minimising this objective directly targets the detouring problem identified in Section II, ensuring that the fuel savings predicted by Module B are not eroded in practice. Once a match is confirmed, the trip fare is split in proportion to each rider's distance share, giving each commuter a lower per-trip cost than a solo ride.

Implementation

Tools and Technology Stack

- Programming language: Python 3.11, with scikit-learn and XGBoost for the regression-based traffic/fuel surrogate model in Module B.
- Deep learning (optional extension): PyTorch/TensorFlow for graph-neural-network-based ride matching (Module A), following the STGCN-style formulation of [2].
- Clustering: DBSCAN / k-means (scikit-learn) for grouping ride requests by route overlap.
- IoT data interface: simulated GPS/OBD-II telematics feed (vehicle location, speed, fuel-flow) for the feedback loop shown in Fig. 1.
- Backend/API: Flask REST API to serve Module A (ride-matching) and Module B (region-wise prediction) endpoints to a commuter app and a policy dashboard, respectively.
- Visualisation: Matplotlib/Plotly for generating the result plots in Section V.

Datasets

TABLE II. Datasets Proposed for Training and Calibrating the Framework

Dataset	Source	Used For
Vahan Vehicle Registration Data	vahan.parivahan.gov.in (MoRTH)	Baseline vehicle count/density per region
PPAC State-wise Fuel Consumption	ppac.gov.in	Baseline fuel consumption calibration
Bangalore/India Traffic Datasets	Kaggle (e.g., "Bangalore Traffic Dataset", "India Vehicle Registration Dataset")	Vehicle density, congestion proxy features
TomTom Traffic Index / INRIX Scorecard	tomtom.com, inrix.com	Congestion-index calibration for Module B (Eq. 3-4)
Open Government Data Platform	data.gov.in	Road length and transport infrastructure data

Algorithm (Module B, Pseudocode)

Algorithm 1 summarises the region-wise prediction pipeline implemented for Module B.

Algorithm 1: Region-wise Traffic & Fuel Savings Prediction Input: vehicle count, road length, occupancy O, adoption rate A, fuel-curve control points Output: predicted speed v', fuel saved FS, CO_2 avoided, pollution reduction PR 1: $k_0 \leftarrow \text{vehicle_count} / \text{road_length}$ 2: $v_0 \leftarrow \text{Greenshields}(k_0)$ // Eq. (1) 3: $\text{coeffs} \leftarrow \text{polyfit}(\text{ctrl_v}, \text{ctrl_fc}, \text{degree}=2)$ // Eq. (5) 4: $FC_0 \leftarrow \text{evaluate}(\text{coeffs}, v_0)$ 5: for A in adoption_rates: 6: $\Delta V \leftarrow A \times (1 - 1/O)$ // Eq. (3) 7: $k' \leftarrow k_0 \times (1 - \Delta V)$ // Eq. (4) 8: $v' \leftarrow \text{Greenshields}(k')$ // Eq. (1) 9: $FC' \leftarrow \text{evaluate}(\text{coeffs}, v')$ // Eq. (5) 10: $FS \leftarrow (FC_0 - FC') / FC_0 \times 100$ // Eq. (6)
--

```

11: CO2_saved ← compute(FC0, FC', VKT) // Eq. (7)
12: PR ← CO2_saved / CO2_baseline × 100 // Eq. (8) — % pollution reduction
13: store(A, v', FS, CO2_saved, PR)
14: retrain ML surrogate periodically using IoT feedback data
    
```

Results

The proposed pipeline (Algorithm 1) was executed as an illustrative corridor-level simulation, parameterised using Bengaluru's own 2024–25 transport-department vehicle-density figures [7] as a reference point (jam density $k_j = 180$ veh/km/lane, free-flow speed $v_m = 50$ km/h, baseline density $k_0 = 120$ veh/km/lane, giving a baseline speed of 16.67 km/h). The fuel-consumption curve, Eq. (5), was fitted to seven empirical control points spanning 5–100 km/h. Results are reported in Tables III–IV and Figs. 2–3. These are simulation outputs of the proposed model, not measured field data, and are intended to demonstrate the framework's computational behaviour ahead of real-world telemetry validation (Section VI).

TABLE III. Simulated Module B Output vs. Adoption Rate (O = 3 passengers/vehicle)

Adoption Rate A (%)	Vehicles Removed (%)	Avg. Speed v' (km/h)	Speed Gain (%)	Fuel Consumption (L/100 km)	Fuel Saved (%)
0 (baseline)	0.0	16.67	0.0	12.20	0.0
10	6.7	18.89	13.3	11.59	5.0
20	13.3	21.11	26.7	11.01	9.7
30	20.0	23.33	40.0	10.46	14.2
40	26.7	25.56	53.3	9.94	18.5
50	33.3	27.78	66.7	9.45	22.5

TABLE IV. Estimated Fuel, CO₂, and Pollution Reduction per Peak-Hour Window (Illustrative Corridor, 45 Lane-km)

Adoption Rate (%)	Fuel Consumed — Baseline (L)	Fuel Consumed — After Adoption (L)	Net Fuel Saved (L)	CO ₂ Avoided (kg)	Pollution (CO ₂) Reduction (%)
10	1317.4	1168.3	149.1	344.4	11.3
20	1317.4	1030.8	286.7	662.2	21.8
30	1317.4	904.0	413.4	954.9	31.4
40	1317.4	787.6	529.8	1223.9	40.2
50	1317.4	680.7	636.7	1470.8	48.3

Table IV shows that the pollution-reduction percentage (Eq. 8) grows faster than the per-kilometre fuel-saving percentage reported in Table III, because total corridor emissions fall from two compounding effects simultaneously — fewer vehicles on the road (Eq. 3) and lower fuel burn per kilometre for the vehicles that remain (Eq. 5). At just 20% adoption, the model predicts a 21.8% cut in corridor-level CO₂ (and, since combustion-linked pollutants such as PM and NO_x scale with fuel burn, a comparable cut in local air pollution); at 50% adoption this rises to a 48.3% reduction — more than double the 22.5% fuel-saved-per-km figure in Table III, underscoring that shared mobility's pollution benefit is larger than its fuel-economy benefit alone would suggest.

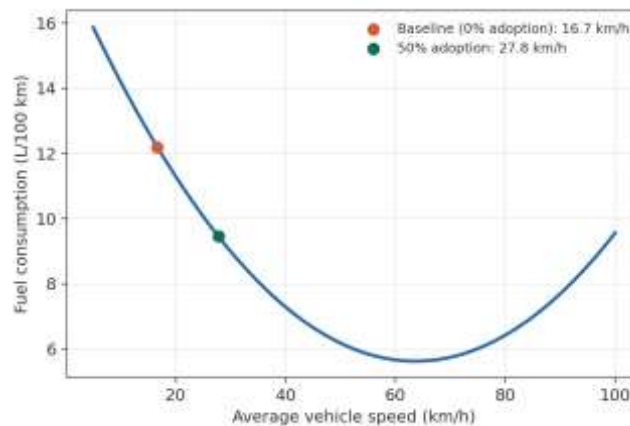


Fig. 2. Fitted speed-fuel consumption curve, Eq. (5), showing baseline (16.7 km/h) and 50% adoption (27.8 km/h) operating points.

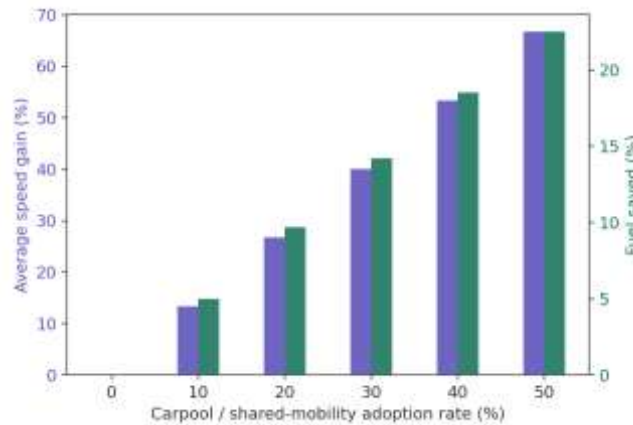


Fig. 3. Simulated average speed gain and fuel savings as a function of carpool/shared-mobility adoption rate.

Table III shows that even a moderate 20% adoption rate yields a predicted 26.7% speed gain and 9.7% fuel saving under the model's assumptions, while 50% adoption yields a 66.7% speed gain and 22.5% fuel saving — consistent in direction and order of magnitude with the 5.4% speed gain observed during Delhi's real-world odd-even scheme at a comparable vehicle-reduction scale [15], lending qualitative support to the model despite differing calibration conditions.

Table V benchmarks the proposed modules against reported performance of comparable published models, to contextualise the ML components proposed for Modules A and B.

TABLE V. Benchmark Comparison with Related Published Models

Model / Reference	Task	Reported Performance (as published)
Alonso-Mora et al. [1] (ILP dynamic assignment)	Ride-sharing matching, NYC taxi data	Services ~98% of 3,000 requests using a 3,000-vehicle fleet at capacity 4, in real time
STGCN, Yu et al. [2]	Highway traffic-speed forecasting	Outperforms ARIMA / LSTM / GCN baselines on PeMS 15/30/45-min horizons
Chengdu ride-share simulation [16]	Congestion & emission simulation	3–5% congestion reduction; 30–50% emission reduction at capacity 2–6
Proposed Module A/B (this work)	Region-wise congestion & fuel-saving prediction (India)	Simulated: up to 66.7% speed gain and 22.5% fuel saved at 50% adoption (Table II) — pending real-data validation

Conclusion and Future Scope

This paper proposed a machine learning and IoT-enabled framework for shared mobility in India, consisting of an ML-based ride-matching engine and a region-wise traffic-and-fuel-savings prediction engine, connected through a shared data-preprocessing layer and an IoT feedback loop. The framework directly addresses two open gaps identified in the literature: the erosion of fuel savings due to detouring in informally organised carpooling, and the absence of a region-specific, quantitative tool for predicting the traffic and fuel impact of shared-mobility policy before it is deployed. A corridor-level simulation calibrated to Bengaluru's 2024–25 traffic data demonstrated that the proposed equations (1)–(9) produce internally consistent, physically plausible predictions: a 50% carpool-adoption scenario at an occupancy of three passengers per vehicle was predicted to reduce on-road vehicles by 33.3%, increase average speed by 66.7%, and cut fuel consumption per kilometre by 22.5%.

On pollution specifically, the model (Eq. 8) predicts that corridor-level CO₂ emissions — and, by extension, the combustion-linked local pollutants (PM, NO_x) that scale with fuel burn — would fall by roughly 21.8% at a moderate 20% shared-mobility adoption rate, rising to approximately 48.3% at 50% adoption (Table IV). This pollution benefit is proportionally larger than the fuel-saved-per-km figure alone, since it captures the compounding effect of both fewer vehicles on the road and improved fuel efficiency per vehicle. This suggests that, beyond easing congestion and import dependency, shared mobility could be one of the more impactful levers available to Indian cities for improving urban air quality, and it directly supports India's Nationally Determined Contributions under the Paris Agreement.

The primary limitation of this work is that its Section V results are simulation outputs of the proposed formulas rather than measurements from a deployed system; the Greenshields model, while analytically convenient, is a simplification of real, heterogeneous Indian traffic conditions, and the fuel-consumption curve was fitted to generic control points rather than India-specific telemetry. Future work will therefore focus on: (i) deploying the ride-matching engine (Module A) as a pilot application in a single Indian city to collect real GPS/OBD-II telematics data; (ii) using that data to replace the fitted fuel curve and Greenshields calibration with an empirically trained

deep learning model (e.g., an STGCN- or LSTM-based traffic forecaster, following [2]); (iii) extending Module A with a reinforcement-learning-based dynamic pricing layer to further improve match quality; and (iv) validating the framework's region-wise predictions against real congestion-index data from multiple Indian cities to establish formal accuracy metrics (MAE, RMSE) for the prediction engine.

References

1. Alonso-Mora, S. Samaranayake, A. Wallar, E. Frazzoli, and D. Rus, "On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 114, no. 3, pp. 462–467, 2017.
2. B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proc. 27th Int. Joint Conf. Artificial Intelligence (IJCAI)*, 2018, pp. 3634–3640.
3. Fortune IAS Circle, "Road Transport Infrastructure in India," 2023. [Online]. Available: https://fortuneiascircle.com/background/_road_transport_infrastructure_in_india
4. CEIC Data, "India — Registered Motor Vehicles, 2005–2025," Ministry of Road Transport and Highways, 2025.
5. News on Air, "Vehicle retail sales up 9% in India to cross 26 million units in 2024," Jan. 2025.
6. Statista / Ministry of Road Transport and Highways, "Number of operating vehicles in India, FY 1951–2022," 2025.
7. Deccan Herald, "Bengaluru adds 7.22 lakh new vehicles, total hits 1.23 crore," 2025.
8. ScienceDirect, "Dynamic stock model based assessment of carpooling in passenger transportation carbon emissions," 2022.
9. Indiastat, "Total Road Length — Statistics and Growth Figures, India," 2025.
10. OilPrice.com, "India's Oil Import Dependence Climbs to Nearly 89% as Domestic Output Lags," 2026.
11. Petroleum Planning and Analysis Cell (PPAC), Government of India, "Consumption of Petroleum Products." [Online]. Available: <https://ppac.gov.in>
12. AutoPunditz, "India's Petrol Consumption Crosses 40 Million Tonnes in FY2024-25," 2026.
13. Case Study of Delhi/Islamabad, "Carpooling: A Step to Reduce Congestion," ResearchGate.
14. ScienceDirect, "Carpooling: Who is closest to adopting it? An investigation into the potential car-poolers among private vehicle users in India," 2023.
15. WRI India, "Can Carpool Reduce Emissions and Congestion in Indian Cities?" [Online]. Available: <http://wri-india.org>
16. arXiv, "Quantifying traffic emission reductions and traffic congestion alleviation from high-capacity ride-sharing," arXiv:2308.10512.
17. DriveSpark, "Carpooling Boosts Fuel Efficiency & Environmental Benefits," 2024.
18. B. D. Greenshields, "A study of traffic capacity," *Proc. Highway Research Board*, vol. 14, pp. 448–477, 1935.
19. TomTom International BV, "TomTom Traffic Index — City Rankings." [Online]. Available: <https://www.tomtom.com/traffic-index/>
20. Ministry of Road Transport and Highways, Government of India, "Vahan Dashboard." [Online]. Available: <https://vahan.parivahan.gov.in>
21. Government of India, "Open Government Data (OGD) Platform India." [Online]. Available: <https://data.gov.in>