AdTrans project, 2023-1-PL01-KA220-HED-000158917

ADTRANS
ADVANCED DATA ANALYSIS FOR TRANSPORTATION AND MOBILITY ACADEMIC TEACHERS

Road Traffic Control: Intelligent Traffic Control through Data-Driven Prediction and Pattern Analysis

Jacek Oskarbski

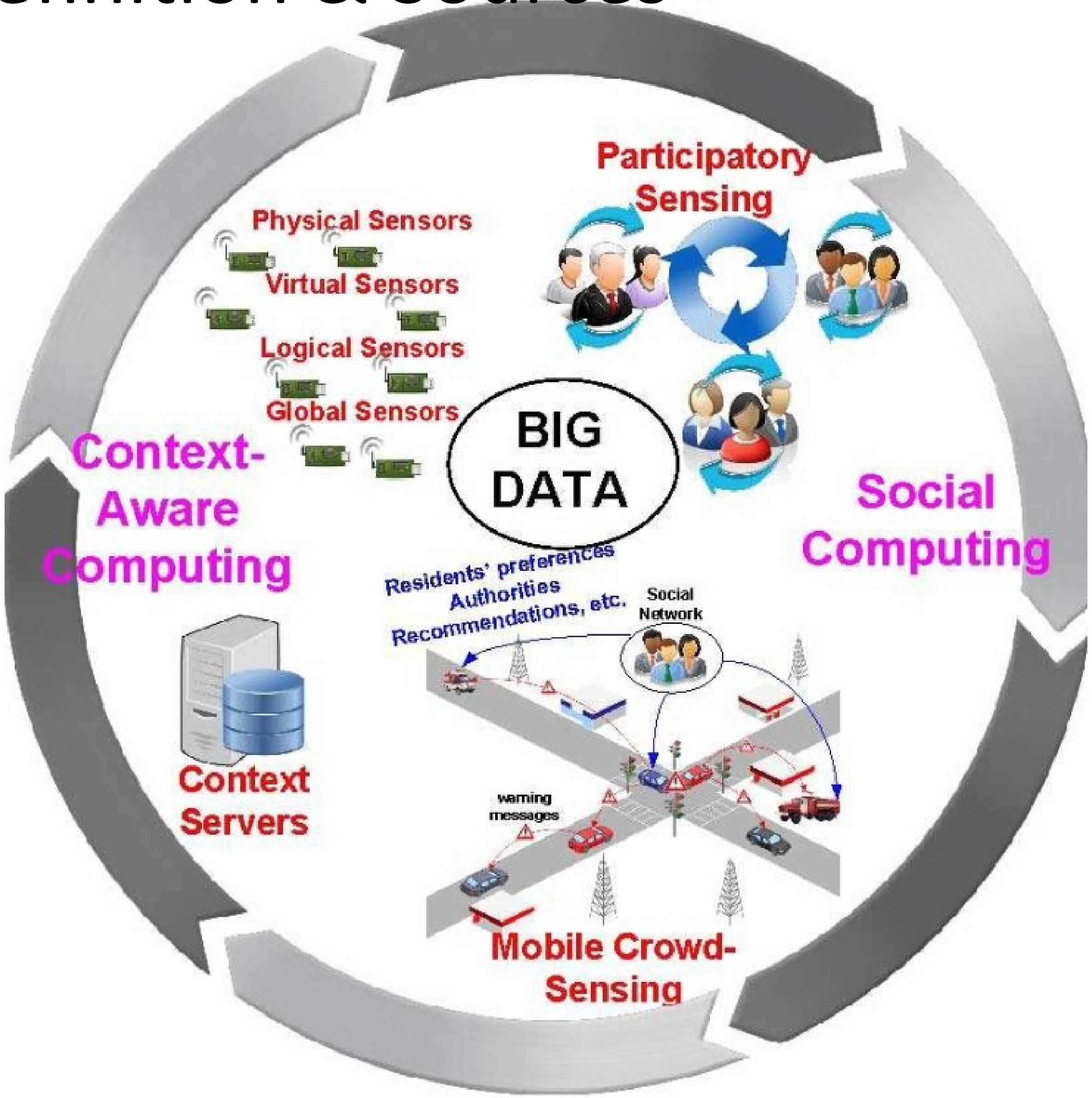
Co-funded by the European Union

Data-Driven Traffic Prediction and Pattern Analysis

- Modern transportation systems generate massive amounts of heterogeneous data (loop detectors, GPS, mobile phones, social media, connected vehicles).
- Big data analytics and machine learning (ML) are transforming traffic management by enabling accurate forecasting of traffic states, early detection of congestion, and optimisation of control strategies.
- Traditional traffic flow models (macroscopic, mesoscopic, microscopic) are being complemented—and in some cases replaced—by data-driven approaches that capture complex, non-linear dynamics of urban mobility.
- **Objectives of the Lecture:**
    - To introduce the concept of big data in transportation and its relevance to intelligent traffic management.
    - To explain machine learning algorithms commonly applied to traffic forecasting (regression, probabilistic models, deep learning).
    - To explore practical applications of predictive modelling in identifying congestion hotspots and peak travel times.
    - To discuss case studies that highlight the impact of data-driven approaches on real-world transport systems.

Big Data in Transportation - Definition & Sources

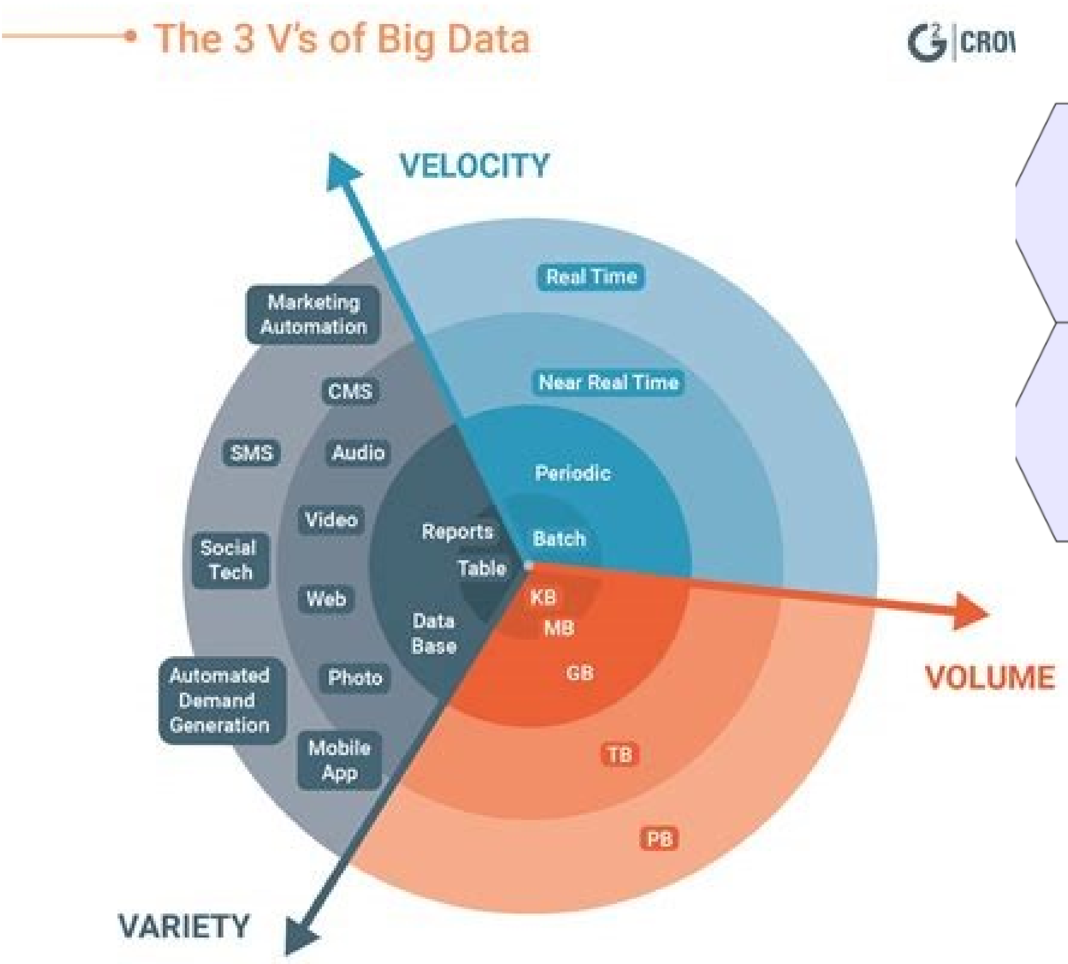
- Big Data in transport refers to large-scale, heterogeneous, high-frequency datasets generated by intelligent transportation systems (ITS) and user devices.
- Main sources include:
  - Infrastructure sensors: inductive loop detectors, radar, thermal and video cameras.
  - Vehicle-based data: GPS traces, in-vehicle sensors, connected vehicles.
  - Mobile and telecom data: anonymised smartphone location data, call detail records.
  - Crowdsourced & social media: applications such as Waze, Google Maps, Twitter providing incident and congestion reports.



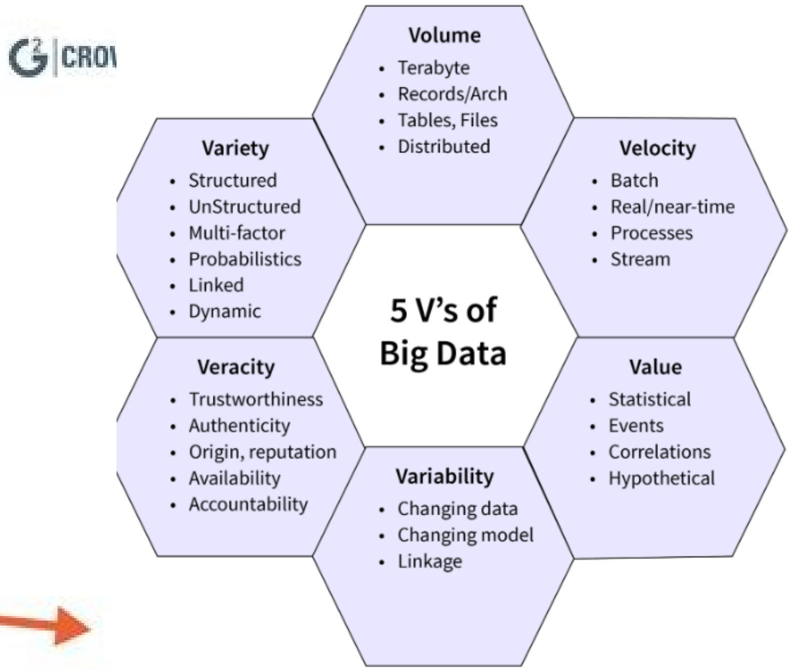
https://vamosarema.com/

Big Data in Transportation - Key Characteristics (3Vs)

- Volume – enormous scale of data (terabytes to petabytes) collected continuously in real-time from millions of devices.
- Velocity – high frequency and speed of data generation (sub-second sensor readings, streaming GPS).
- Variety – diverse data formats: structured (traffic counts), semi-structured (GPS traces), unstructured (social media text, images).
- Some authors add Veracity – uncertainty/noise in data – and Value – usefulness of extracting actionable insights.



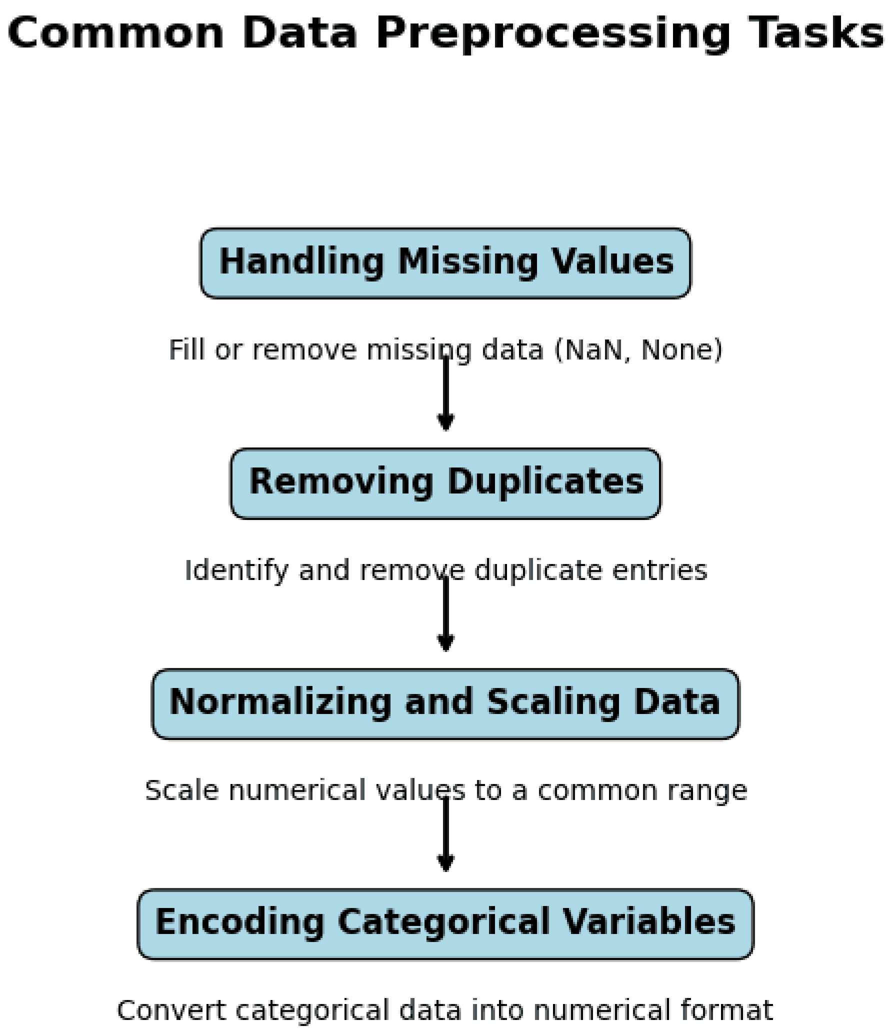
https://www.g2.com/articles/big-data



https://www.scaler.com/topics/5-vs-of-big-data

Big Data in Transportation - Importance of Data Quality

- Missing values: detector failures, network communication gaps.
- Bias in data: unequal distribution (e.g., smartphone data more common in urban than rural areas).
- Noise & anomalies: errors due to weather, sensor miscalibration, or malicious reporting.
- Data cleaning & preprocessing are critical for reliable predictive modelling.



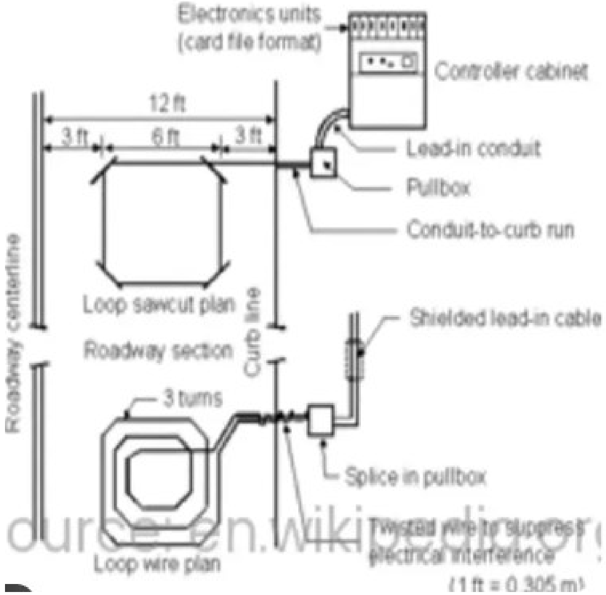
https://emitechlogic.com/how-to-automate-data-preprocessing-in-python

Traffic Data Collection Methods - Inductive Loop Detectors (ILDs)

- Embedded beneath the road surface; measure vehicle presence, count, and occupancy.
- Provide high temporal resolution data but limited spatial coverage.
- Advantages: mature technology, widely deployed in highways and urban arterials.
- Limitations: costly installation/maintenance, not suitable for flexible/mobile applications.

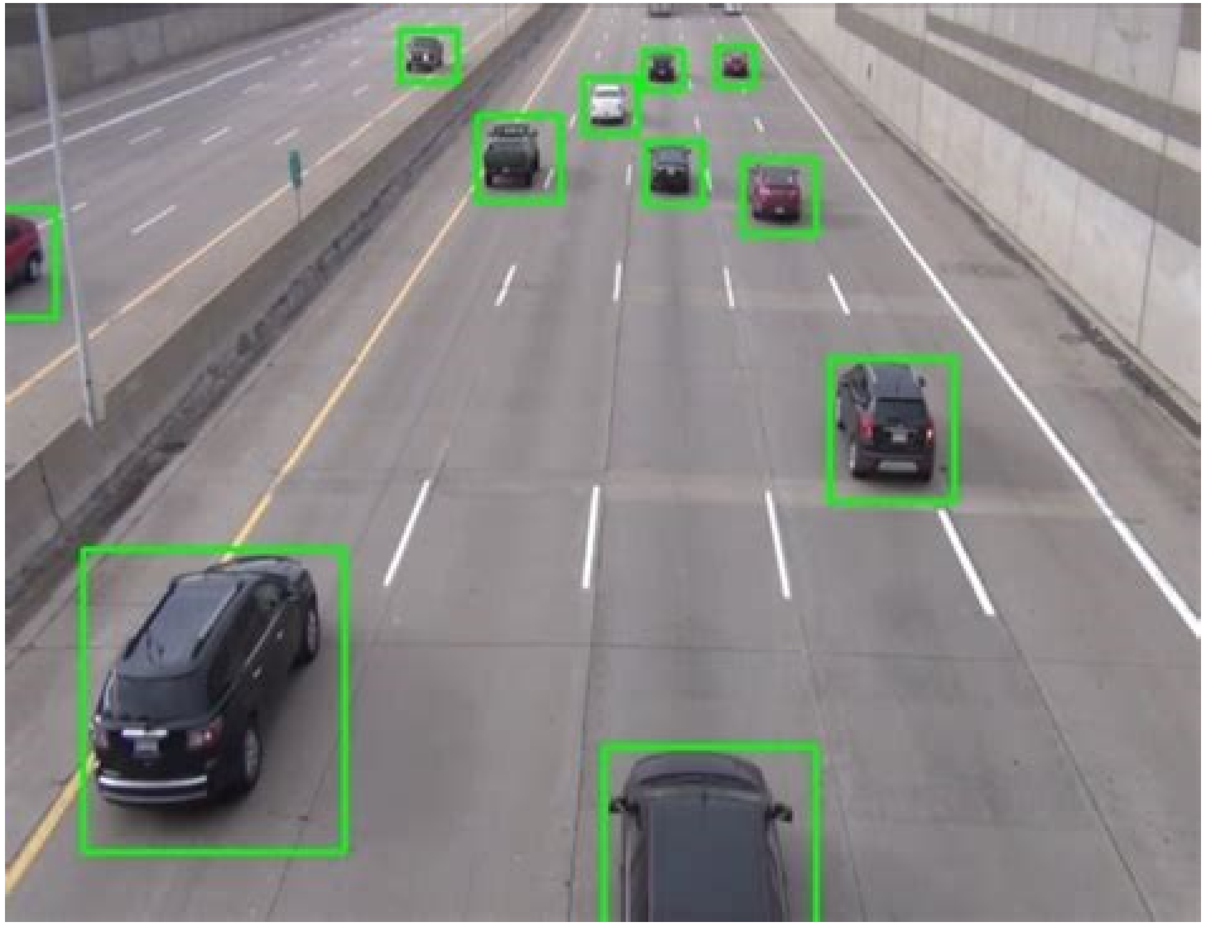
Inductive Loop Detectors

- Inductive Loop Detectors (ILDs) are essential in modern transportation for detecting vehicle movement
- They work on electromagnetic induction with a wire loop embedded in roads connected to a detector
- The two main components are the loop and the detector
- While ILDs are reliable, newer detection systems like video and lidar are emerging



Traffic Data Collection Methods - Video Cameras & Image Recognition

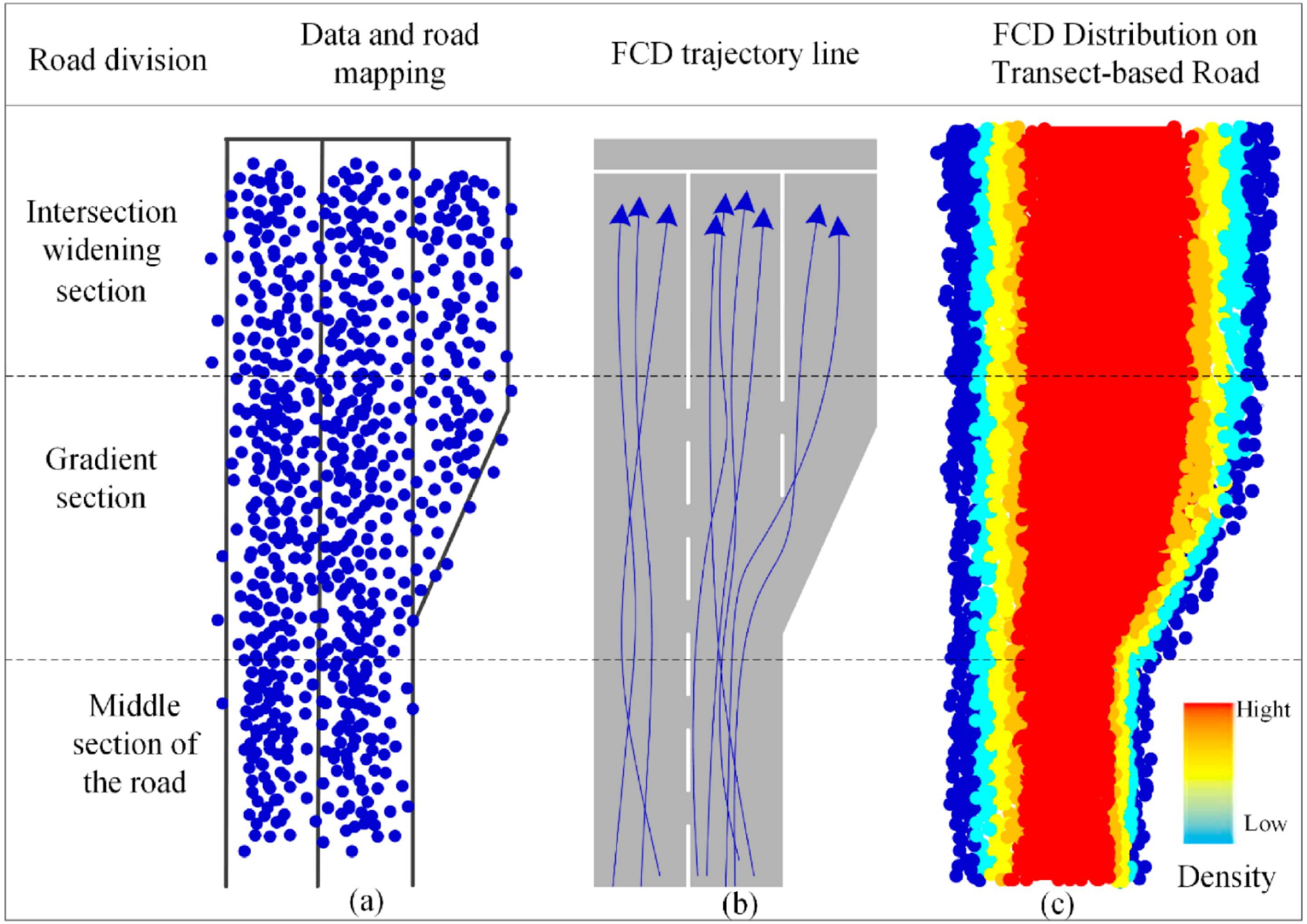
- Capture continuous streams of traffic images and video.
- Computer vision (CV) algorithms and AI/ML detect vehicles, classify types, estimate speeds.
- Applications: traffic counts, queue length estimation, incident detection, lane discipline monitoring.
- Challenges: sensitive to weather (fog, rain, snow) and lighting conditions.



Bounding-box and annotation" infographic, featured on Infizoom

Traffic Data Collection Methods - Floating Car Data (FCD – GPS / Smartphones)

- Data from GPS-equipped vehicles and smartphones provide real-time trajectories.
- Applications: average speed estimation, travel time reliability, route choice behaviour.
- Advantages: wide coverage, low infrastructure cost, real-time updates.
- Issues: sampling bias (not all vehicles equipped), privacy concerns, dependence on telecom networks.



Li, X.; Wu, Y.; Tan, Y.; Cheng, P.; Wu, J.; Wang, Y. Method Based on Floating Car Data and Gradient-Boosted Decision Tree Classification for the Detection of Auxiliary Through Lanes at Intersections. ISPRS Int. J. Geo-Inf. 2018, 7, 317. https://doi.org/10.3390/ijgi7080317

Traffic Data Collection Methods - Crowdsourced & Social Media Data

- Platforms like Waze, Google Maps, TomTom collect user-reported incidents, congestion, road hazards.
- Social media (Twitter, Facebook) adds real-time information on disruptions (accidents, weather, strikes).
- Benefits: high responsiveness, wide participation, complements sensor data.
- Risks: data veracity (false reports, malicious inputs), uneven coverage across demographics and regions.

https://fnews.ai/google-maps-and-waze-updated-with-enhanced-incident-reporting-and-destination-guidance

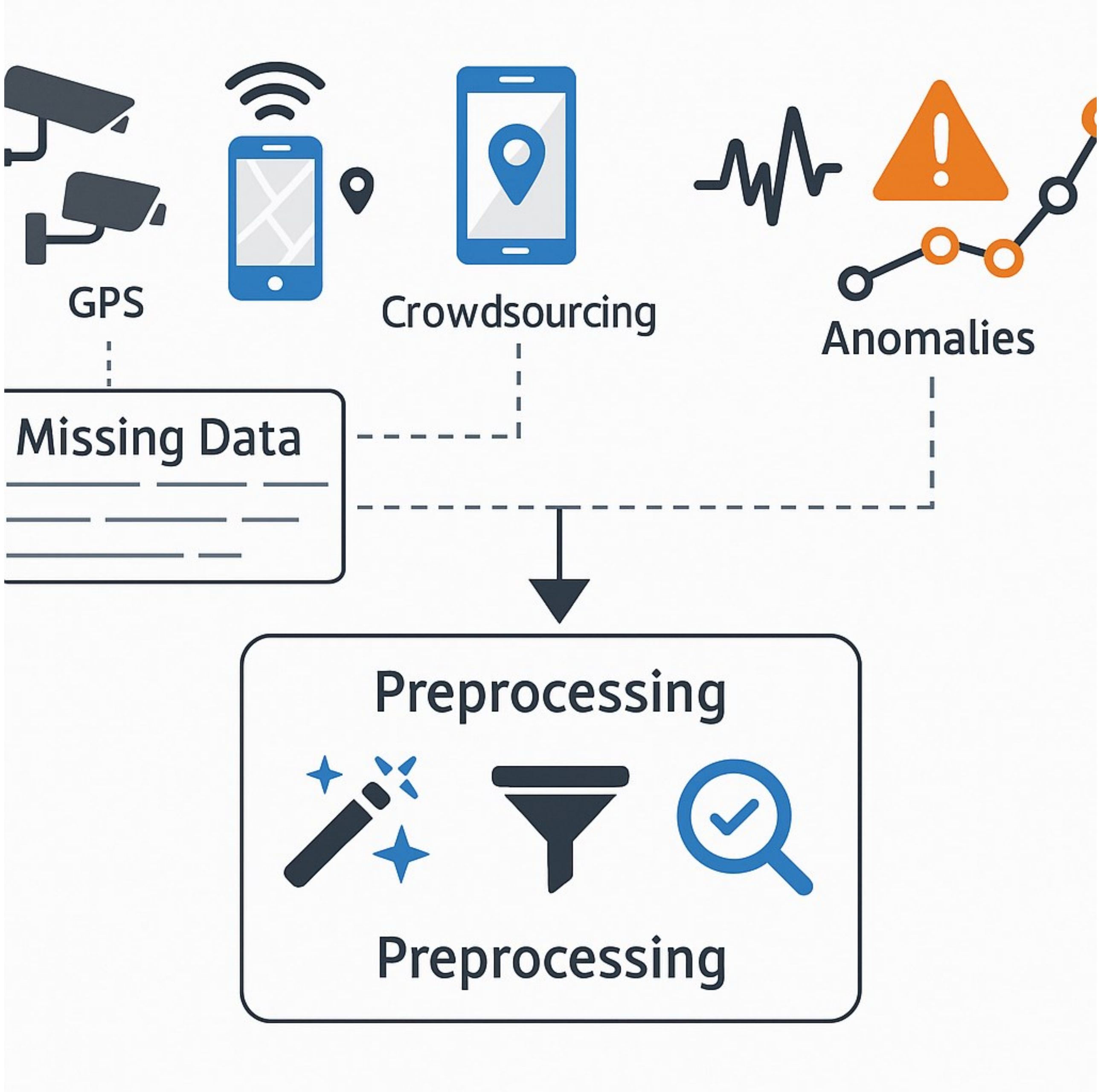
Challenges in Data Processing - Heterogeneity of Data Sources

- Data comes from diverse systems: inductive loops, GPS traces, cameras, mobile apps, crowdsourced platforms.
- Different formats: structured (traffic counts), semi-structured (GPS trajectories), unstructured (tweets, images).
- Integration requires advanced data fusion techniques to align spatial and temporal resolutions.

https://www.smatstraffic.com/blog/5-reasons-floating-car-data

Challenges in Data Processing - Missing Data, Noise, and Anomalies

- Missing values caused by detector malfunctions, communication failures, or incomplete coverage.
- Noise and anomalies introduced by sensor miscalibration, harsh weather (fog, snow, rain), or malicious reporting.
- Preprocessing steps (imputation, filtering, anomaly detection) are critical to ensure model reliability.

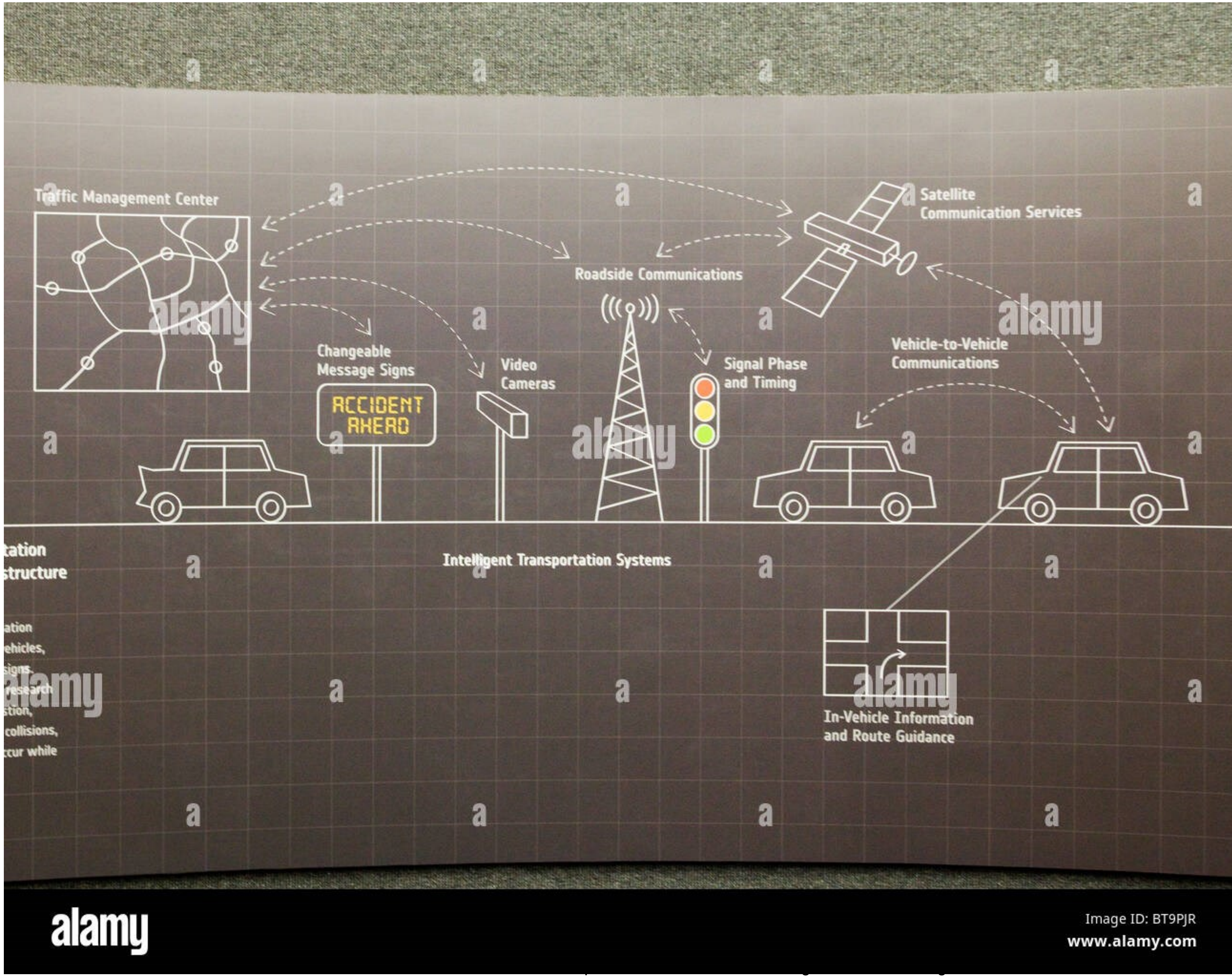


Challenges in Data Processing - Privacy and Data Security Issues

- Location data from GPS and smartphones may expose personal mobility patterns.
- Need for anonymisation and encryption in data storage and exchange.
- Legal frameworks: GDPR (EU), CCPA (California) regulate data use and sharing.
- Balancing data utility vs. privacy protection remains a major challenge.

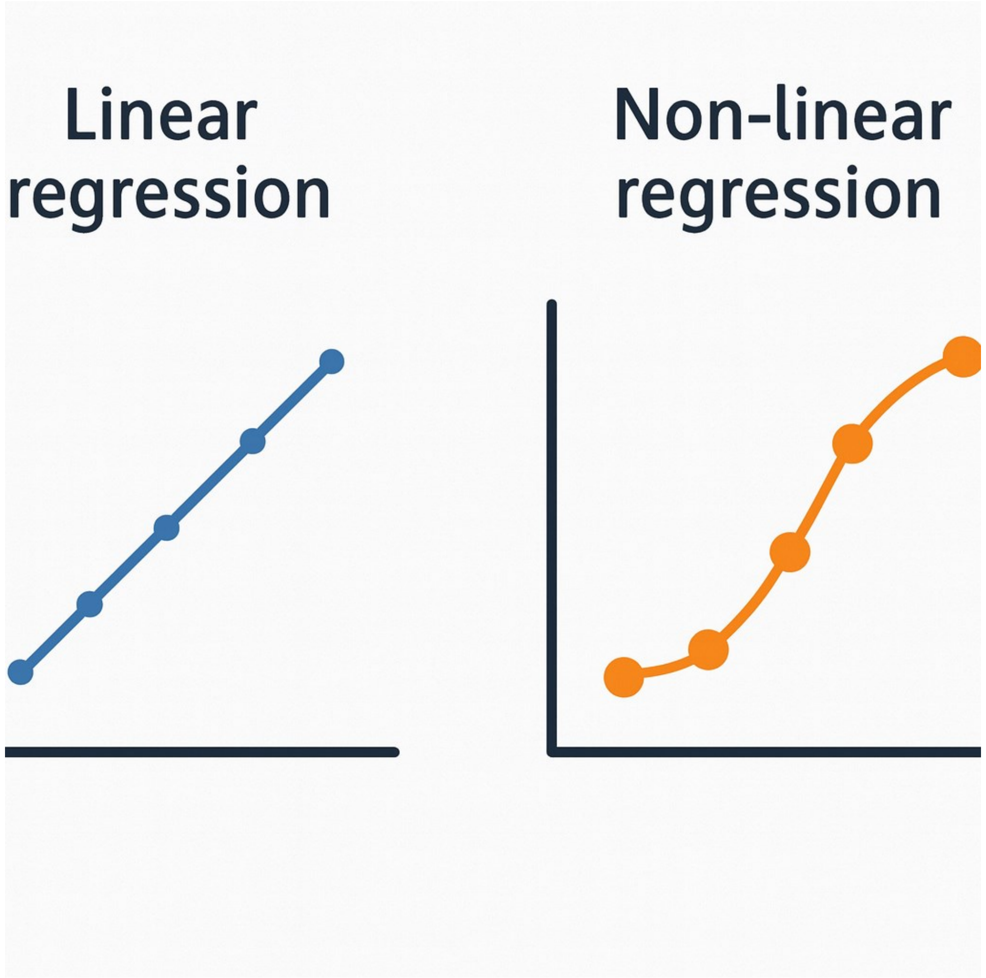
Challenges in Data Processing - Need for Standardisation

- Lack of common data formats and exchange protocols across agencies/vendors.
- Standards like DATEX II (Europe) or NTCIP (US) support interoperability but adoption is uneven.
- Without harmonisation, integration into Intelligent Transport Systems (ITS) is fragmented and costly.



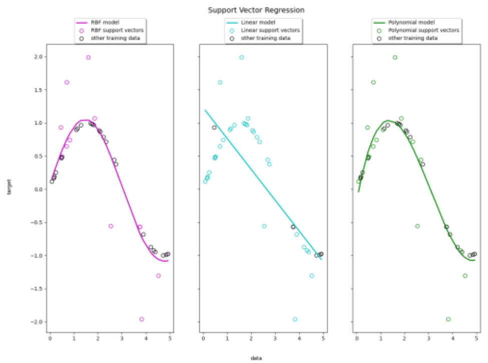
Machine Learning Basics for Traffic Forecasting - Linear and Non-linear Regression

- Linear regression: Establishes a straight-line relationship between traffic variables (e.g., flow vs. time). Useful for short-term forecasting in stable conditions.
- Non-linear regression: Captures complex traffic phenomena such as saturation, congestion onset, shockwaves.
- Example: Polynomial regression, Support Vector Regression (SVR).



Machine Learning Basics for Traffic Forecasting - Linear and Non-linear Regression

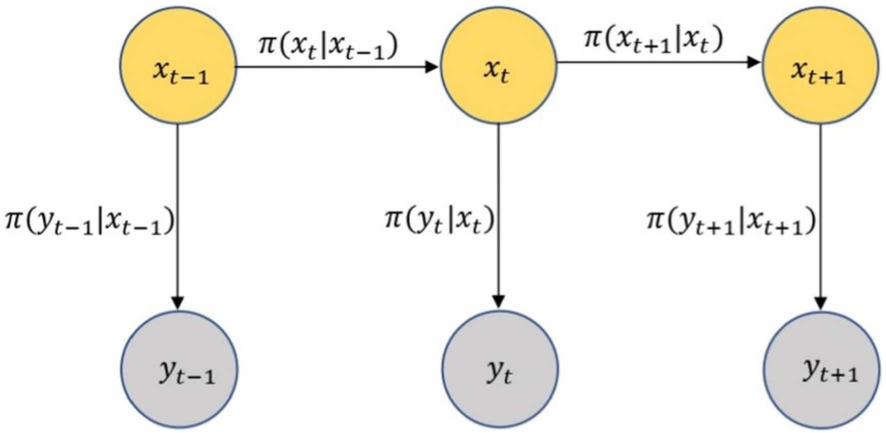
- Polynomial Regression:
  - Extends linear regression by including polynomial terms (e.g. , $y = \beta_0 + \beta_1 x + \beta_2 x^2 + ..$
  - Captures non-linear trends in traffic flow (e.g., sudden congestion growth after a threshold of vehicle density).
  - Example: predicting speed–density relationships where flow increases with density up to a critical point, then decreases (fundamental diagram).
  - Advantage: simple, interpretable; works for small datasets.
  - Limitation: may overfit and fail with high-dimensional data.

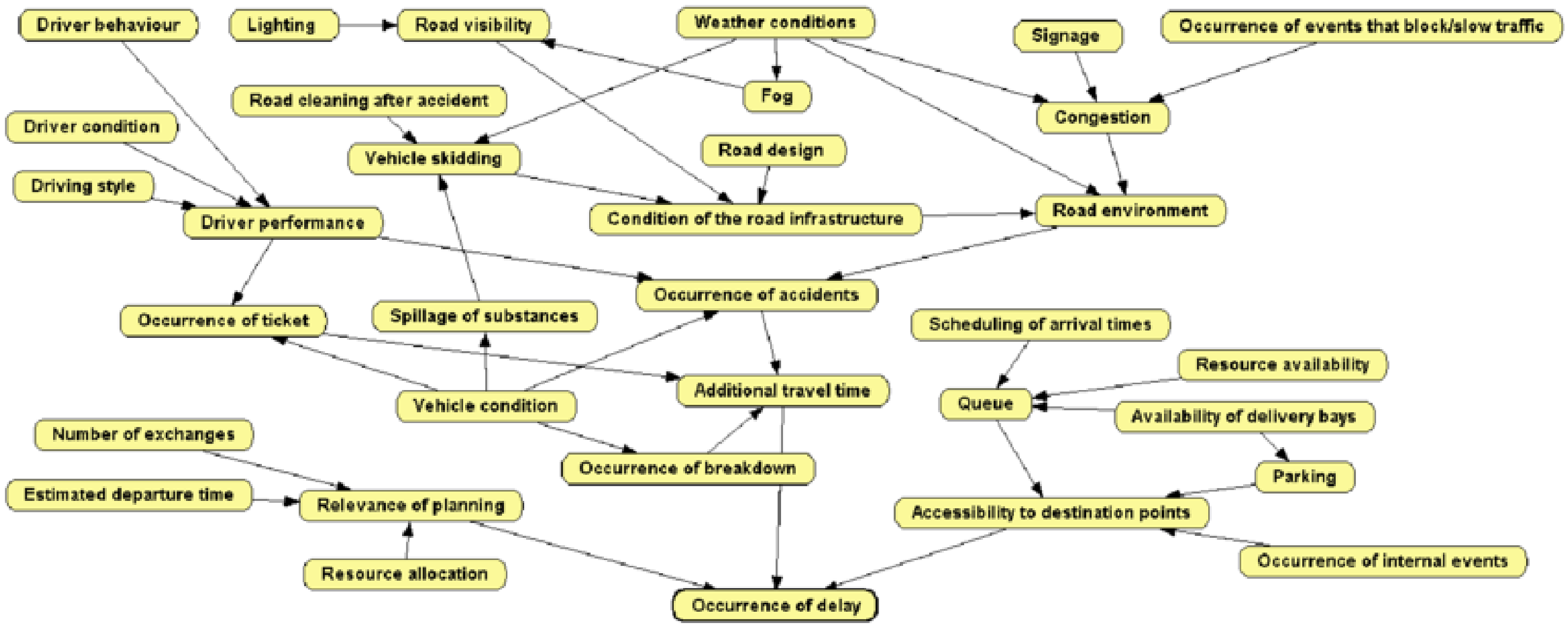


- Support Vector Regression (SVR)
  - Based on Support Vector Machines, adapted for regression tasks.
  - Finds a regression line (or hyperplane) that best fits the data within a margin of tolerance (epsilon-insensitive loss).
  - Handles non-linear patterns through kernel functions (e.g., radial basis function).
  - Widely used for short-term traffic flow forecasting, outperforming linear models in capturing variability.
  - Advantage: robust against noise, flexible.
  - Limitation: computationally expensive for very large datasets.

Machine Learning Basics for Traffic Forecasting - Probabilistic Models

- Hidden Markov Models (HMM): Capture time-series state transitions, e.g., from “free flow” → “congested” → “recovery”.
- Bayesian Networks: Encode causal relationships (e.g., incident → congestion → travel time delay).
- Strength: handle uncertainty and missing data.
- Application: incident detection, travel time reliability estimation

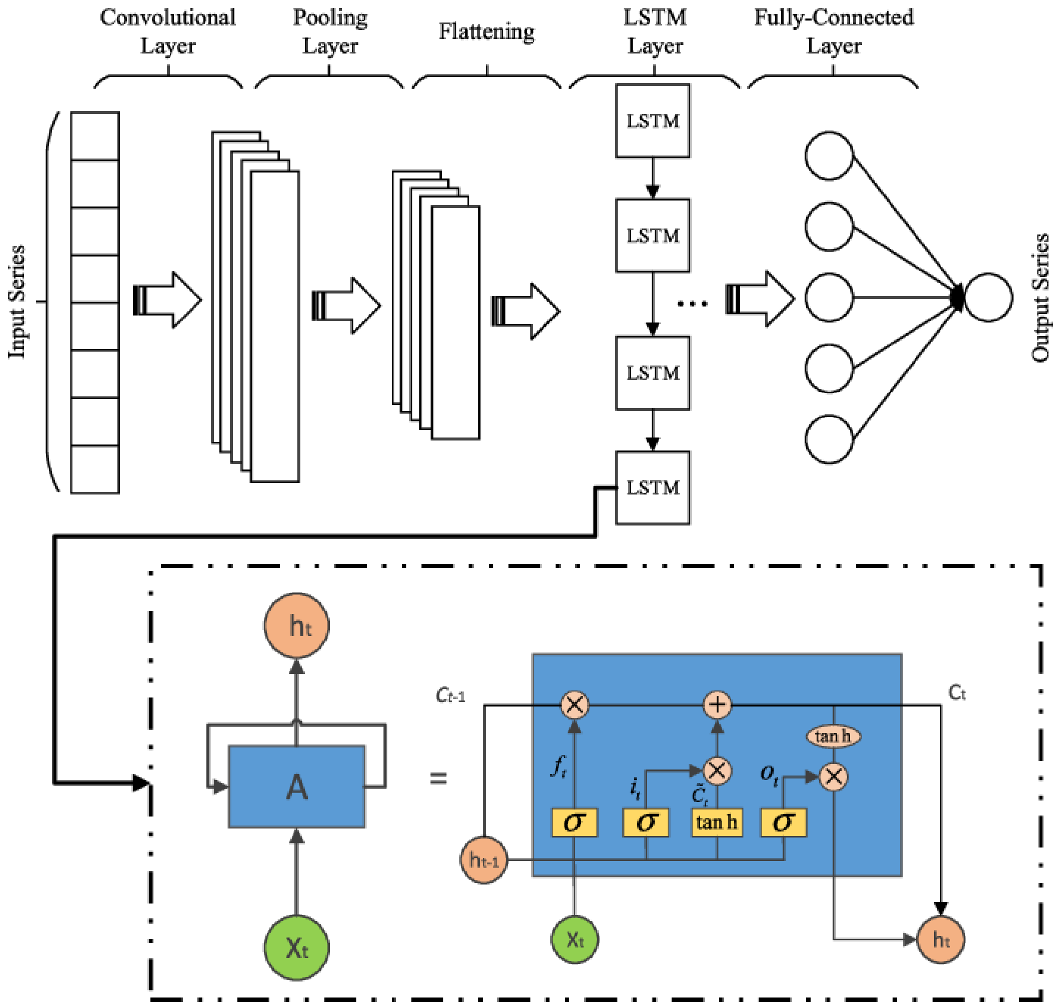




el Bouhadi, Ouafae & Azmani, Monir & Azmani, Abdellah & ftouh, Mouna. (2022). Using a Fuzzy-Bayesian Approach for Predictive Analysis of Delivery Delay Risk. International Journal of Advanced Computer Science and Applications. 13. 10.14569/IJACSA.2022.0130740.

Machine Learning Basics for Traffic Forecasting - Classical Algorithms vs. Deep Learning Approaches

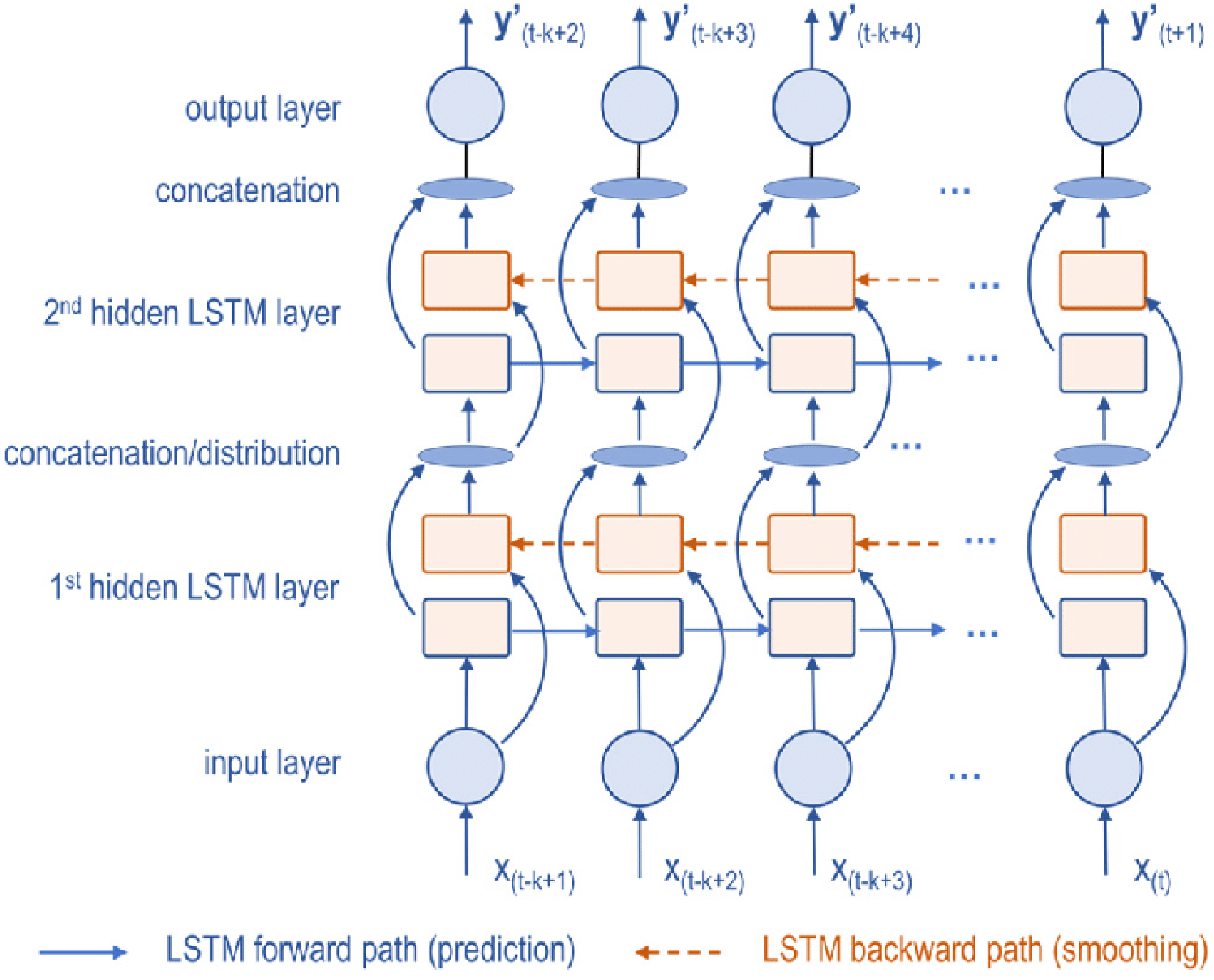
- Classical ML methods: Decision trees, Random Forests, k-NN, SVM.
  - Strength: interpretable, efficient for smaller datasets.
  - Limitation: struggle with high-dimensional, spatio-temporal data.
- Deep learning models:
  - RNN / LSTM: capture temporal dependencies in sequential data.
  - CNN: used for spatial correlations in traffic networks or image-based traffic data.
  - GNN: model complex road network topologies.
  - Strength: excellent at modelling non-linear, high-dimensional data.
  - Limitation: require large labelled datasets, less interpretable ("black box").



Deeplearning Structure Of 1dcnn And Lstm. https://dietmake.coesca.com/what-is-the-difference-between-cnn-and-lstm-time-series.html

Neural Networks and Deep Learning - Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) for Time-Series Forecasting

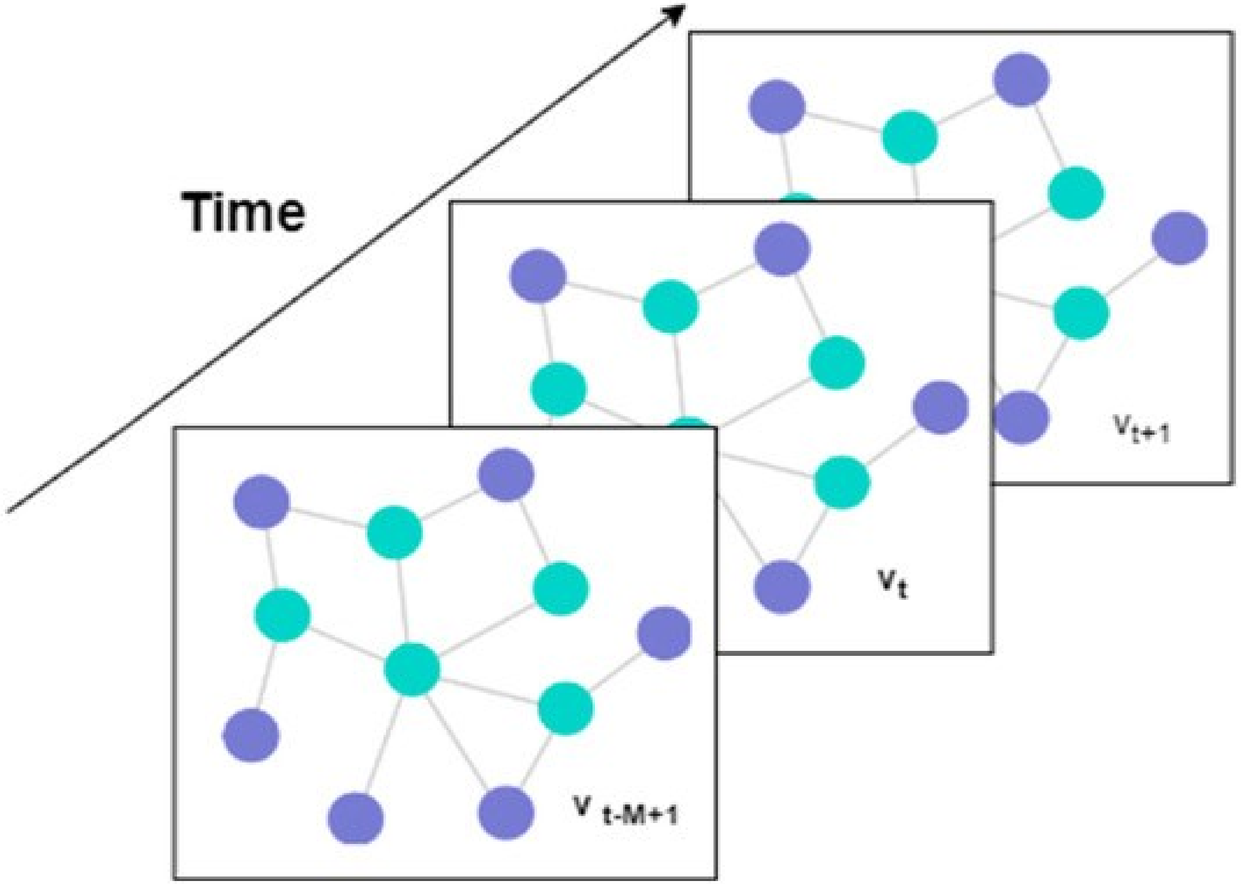
- RNNs are designed to model sequential dependencies, making them suitable for traffic flow and travel time prediction.
- Challenge in basic RNNs: vanishing gradient problem, limiting their ability to learn long-term dependencies.
- LSTM networks overcome this limitation with memory cells and gates (input, forget, output) that capture long-range temporal patterns.
- Applications:
  - Short-term traffic volume prediction (e.g., next 5–15 min).
  - Travel time forecasting along freeway corridors.



LSTM architecture schematic illustrating forward and backward propagation through time. Source: "Schematic Of The Lstm Model Download Scientific Diagr," vrogue.co

Neural Networks and Deep Learning - Graph Neural Networks (GNN) for Road Networks

- Road networks can be represented as graphs (nodes = intersections, edges = road links).
- GNNs learn spatial correlations by propagating information across connected nodes.
- Particularly powerful when combined with temporal models (e.g., ST-GCN: Spatio-Temporal Graph Convolutional Networks).
- Applications:
  - Network-wide traffic flow prediction.
  - Identifying congestion propagation across a road network.
- Example: ST-GCN models both time dependencies (temporal graph) and network structure (spatial graph)



Singh, V.; Sahana, S.K.; Bhattacharjee, V. Integrated Spatio-Temporal Graph Neural Network for Traffic Forecasting. Appl. Sci. 2024, 14, 11477. https://doi.org/10.3390/app142411477

Neural Networks and Deep Learning - Convolutional Neural Networks (CNN) for Video/Image-Based Traffic Analysis

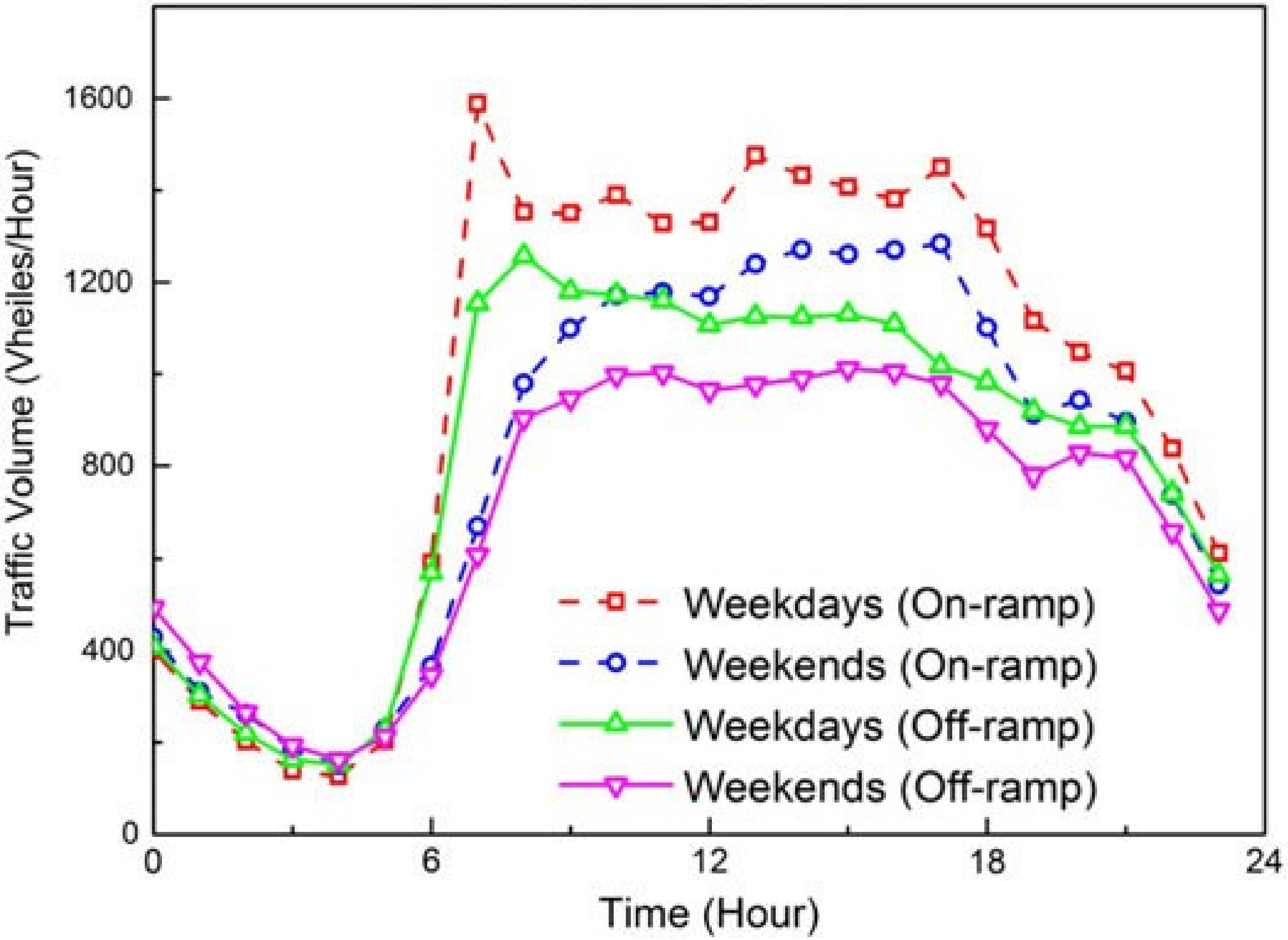
- CNNs excel at image recognition tasks by extracting hierarchical spatial features.
- In transport, they are applied to video feeds from roadside cameras for:
  - Vehicle detection & classification (cars, trucks, buses, motorcycles).
  - Speed estimation using frame-to-frame displacement.
  - Incident detection (stalled vehicles, crashes).
  - Pedestrian and cyclist recognition for safety studies.
- Limitation: performance decreases under adverse weather/lighting, mitigated by thermal or infrared imaging.

(a) (b) (c) (d)

Zhang, F.; Li, C.; Yang, F. Vehicle Detection in Urban Traffic Surveillance Images Based on Convolutional Neural Networks with Feature Concatenation. Sensors 2019, 19, 594. https://doi.org/10.3390/s19030594

Traffic Pattern Analysis - Identification of Daily and Weekly Peaks

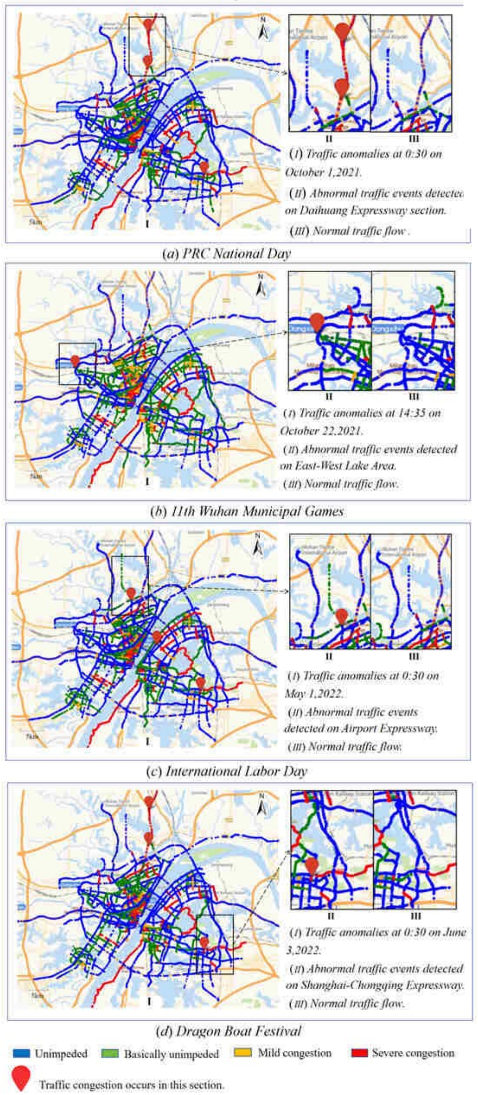
- Traffic demand exhibits strong periodicity:
  - Daily cycles: Morning and evening rush hours (commuting patterns).
  - Weekly cycles: Higher weekday traffic vs. reduced weekend volumes.
- Machine learning methods can automatically extract such periodic patterns from large datasets (GPS traces, loop detectors).
- Applications:
  - Optimization of traffic signal timings.
  - Public transport scheduling aligned with peak demand.



He, Hong-di & Wang, Jun-li & Wei, Hai-rui & Ye, Cheng & Ding, Yi. (2015). Fractal behavior of traffic volume on urban expressway through adaptive fractal analysis. Physica A: Statistical Mechanics and its Applications. 443. 10.1016/j.physa.2015.10.004.

Traffic Pattern Analysis - Anomaly Detection (Accidents, Weather Events, Special Events)

- Accidents: Cause sudden, localized disruptions detectable as sharp drops in speed or unexpected queues.
- Weather events: Rain, snow, fog alter both traffic flow capacity and driver behavior.
- Special events: Stadium matches, concerts, strikes create atypical demand surges.
- ML techniques (e.g., clustering, autoencoders) distinguish between recurrent patterns and non-recurrent anomalies.
- Applications:
  - Early incident detection.
  - Dynamic rerouting and traveler information.



Mao, Y.; Shi, Y.; Lu, B. Detecting Urban Traffic Anomalies Using Traffic-Monitoring Data. ISPRS Int. J. Geo-Inf. 2024, 13, 351. https://doi.org/10.3390/ijgi13100351

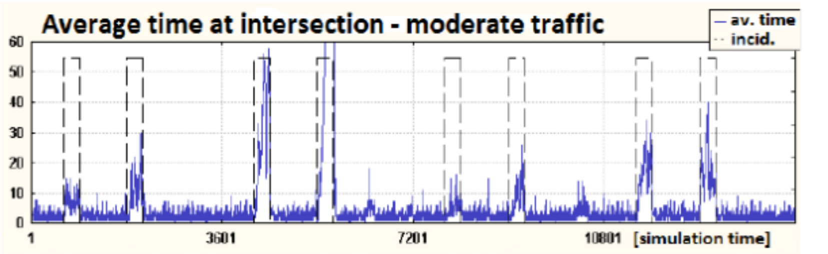
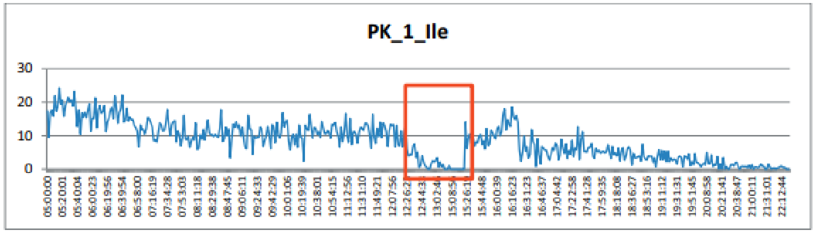
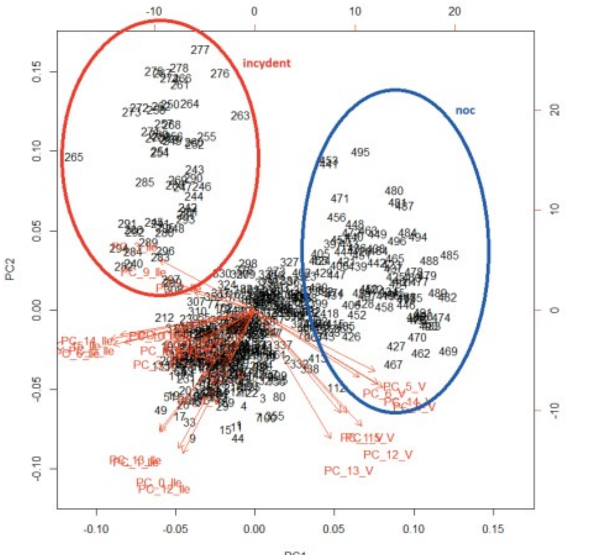


Fig. 5. Comparison of average occupancy indication with simulation times of incidents.

Oskarbski, Jacek & Zawisza, Marcin & Żarski, Karol. (2016). Automatic Incident Detection at Intersections with Use of Telematics. Transportation Research Procedia. 14. 3466-3475. 10.1016/j.trpro.2016.05.309.

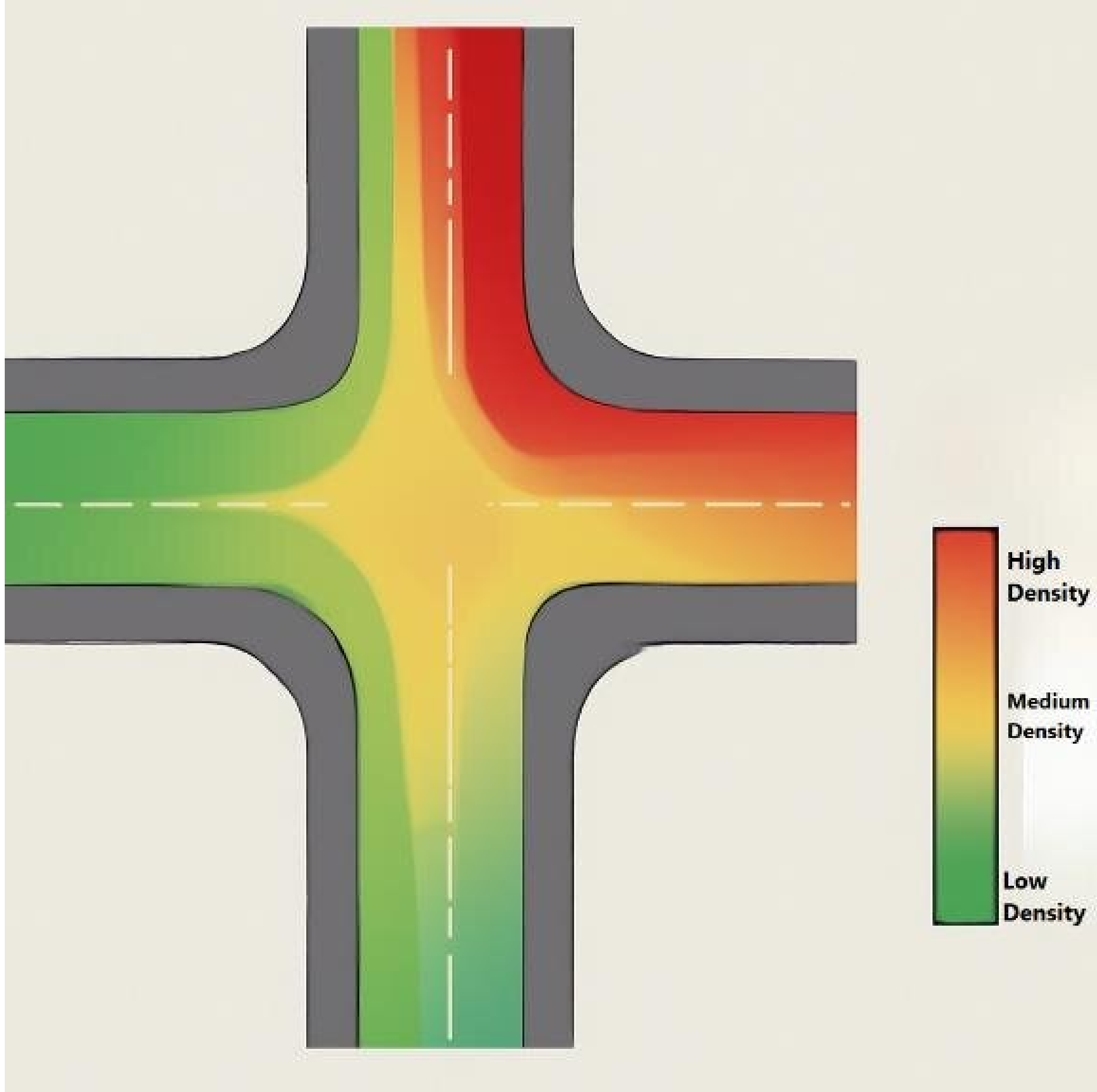




Oskarbski J, Palikowska K, Żarski K. Systemy automatycznego wykrywania zdarzeń niepożądanych w miastach. *Prace Naukowe Politechniki Warszawskiej - Transport.* (2016);114(null):245-254.

Traffic Pattern Analysis - Heatmaps of Traffic Density and Flow

- Visual analytics tool: Represents spatial distribution of traffic volumes/speeds across networks.
- Heatmaps enable intuitive identification of congestion hotspots, bottlenecks, and flow dynamics.
- When combined with temporal layers (e.g., hourly maps), they provide spatio-temporal insights.
- Applications:
  - Urban mobility dashboards for city authorities.
  - Policy evaluation (e.g., effect of congestion pricing or infrastructure changes).

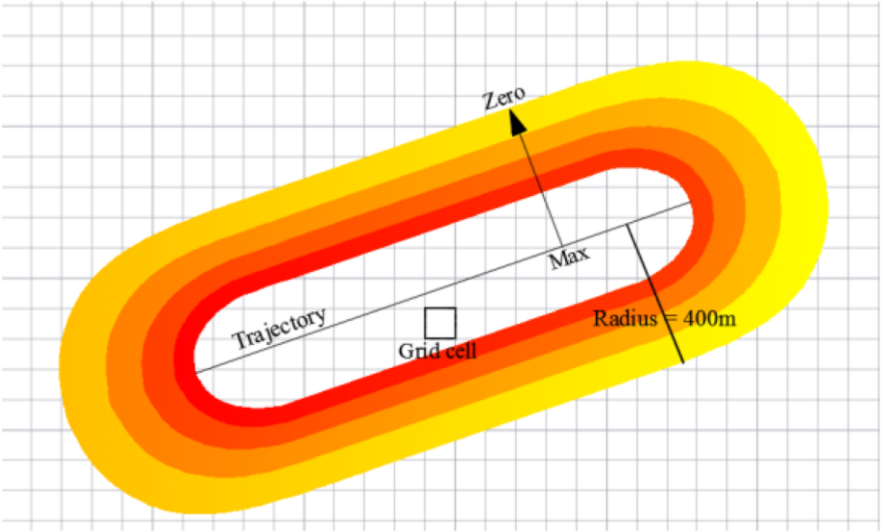


Khaleefah, Rafid. (2025). An Intelligent Traffic Control System using Edge-Cloud Computing and Deep Learning for Smart Cities. International Journal of Engineering Research and. 14. IJERTV14IS070169. 10.17577/IJERTV14IS070169.

Real-Time Prediction in Large Cities

- Megacities such as Los Angeles and Beijing have pioneered real-time traffic forecasting systems to mitigate chronic congestion.
- Data-driven platforms integrate massive data streams from thousands of fixed and mobile sensors.
- Achievements include:
    - Los Angeles: real-time predictive analytics integrated into the Automated Traffic Surveillance and Control (ATSAC) system.
    - Beijing: large-scale deployment of floating car data (GPS from taxis) and video-based monitoring for urban arterials.

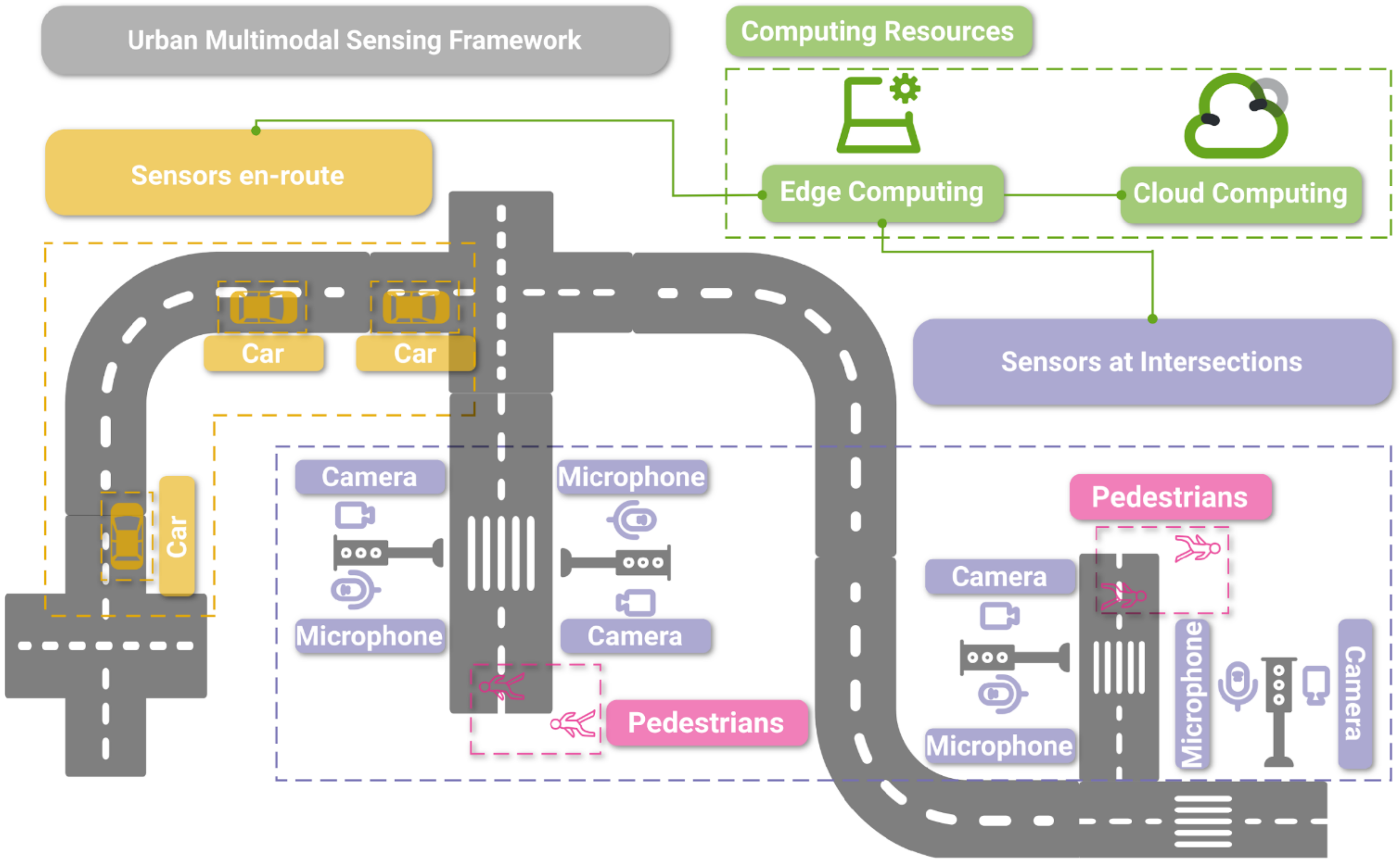
Inside the ATSAC control room where staff monitor live traffic feeds from hundreds of cameras across Los Angeles.
Source: *KCRW 89.9 FM, Photo: Gideon Brower*



Zhang, J.; Chen, F.; Wang, Z.; Wang, R.; Shi, S. Spatiotemporal Patterns of Carbon Emissions and Taxi Travel Using GPS Data in Beijing. *Energies* **2018**, *11*, 500. https://doi.org/10.3390/en11030500

Real-Time Prediction in Large Cities - Integration of GPS and Camera Data

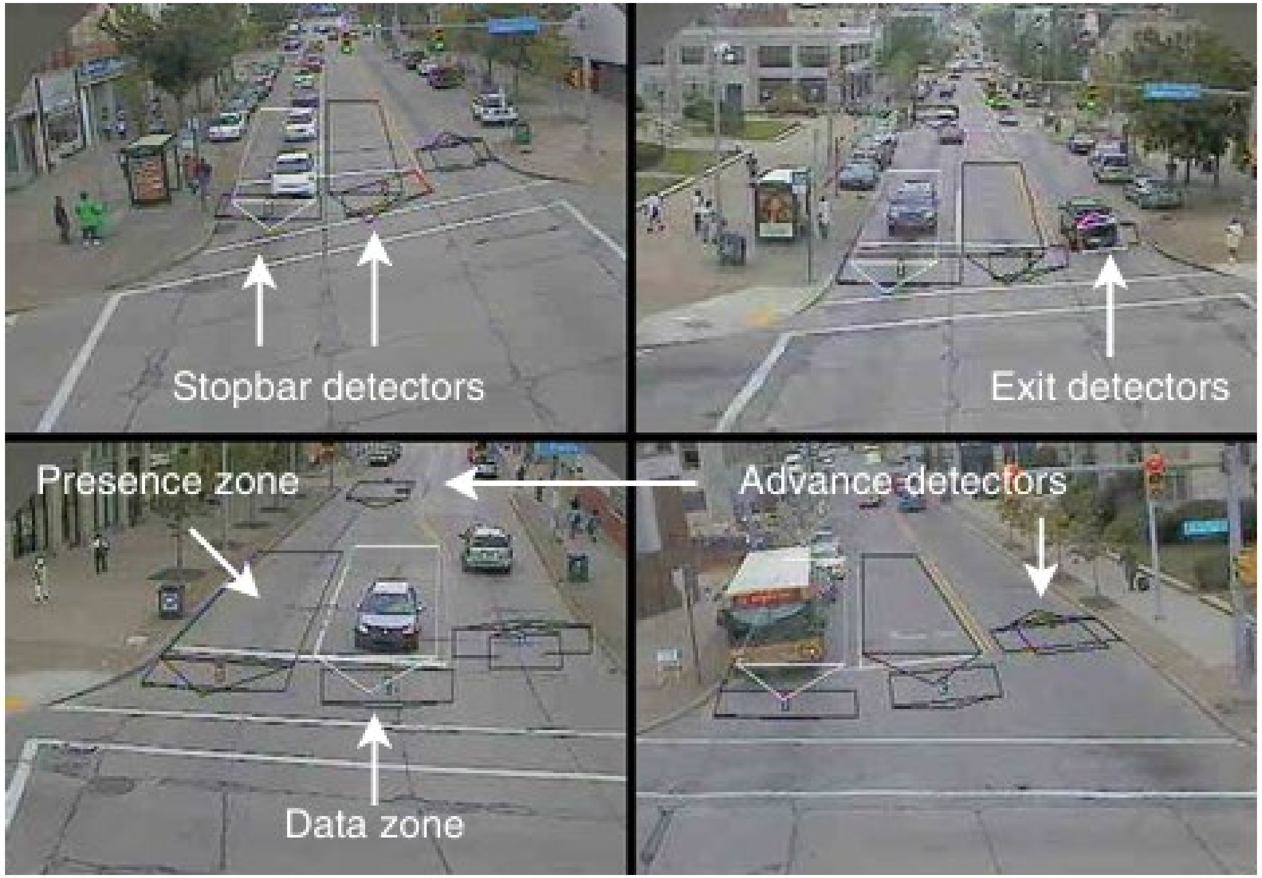
- GPS trajectories from vehicles and smartphones capture macroscopic flow patterns and route choice dynamics.
- Camera-based monitoring provides microscopic details: queue length, vehicle classification, incidents.
- Fusion of these complementary sources improves the spatio-temporal coverage of prediction models.
- Machine learning models (e.g., LSTM, ST-GCN) exploit this heterogeneous data to deliver robust short-term forecasts.



Piadyk, Y.; Rulff, J.; Brewer, E.; Hosseini, M.; Ozbay, K.; Sankaradas, M.; Chakradhar, S.; Silva, C. StreetAware: A High-Resolution Synchronized Multimodal Urban Scene Dataset. Sensors 2023, 23, 3710. https://doi.org/10.3390/s23073710

Real-Time Prediction in Large Cities - Improved Traffic Signal Management

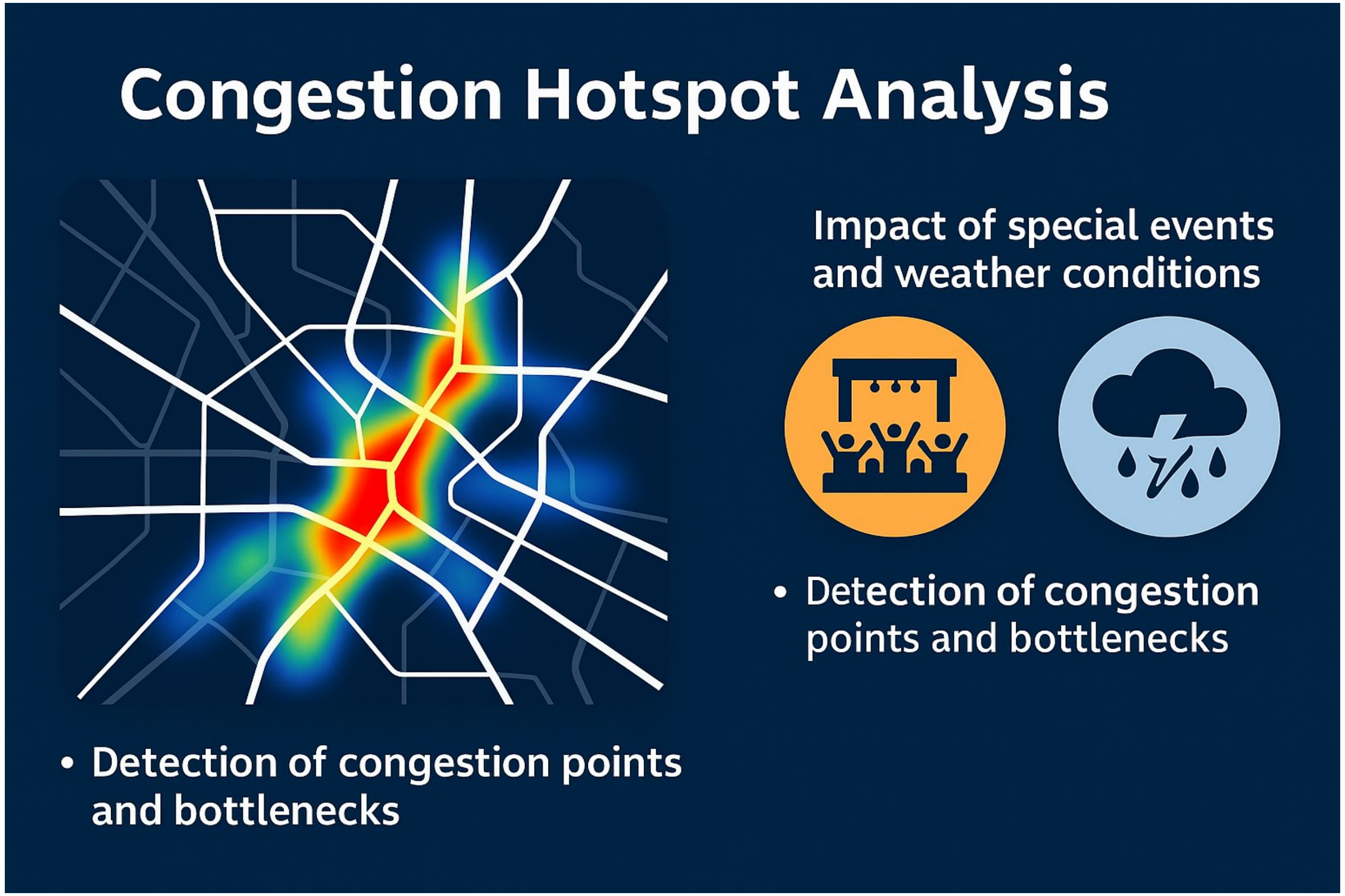
- Real-time prediction feeds directly into adaptive traffic signal control systems.
- Example applications:
  - Preemptive green extensions to mitigate incoming platoons.
  - Dynamic offset adjustments for corridors.
  - Priority strategies for public transport.
- Reported benefits:
  - 10–20% reduction in average delay at intersections.
  - Increased reliability of travel times during peak hours.



SURTRAC:Scalable Urban Traffic Control Stephen F. Smith and others

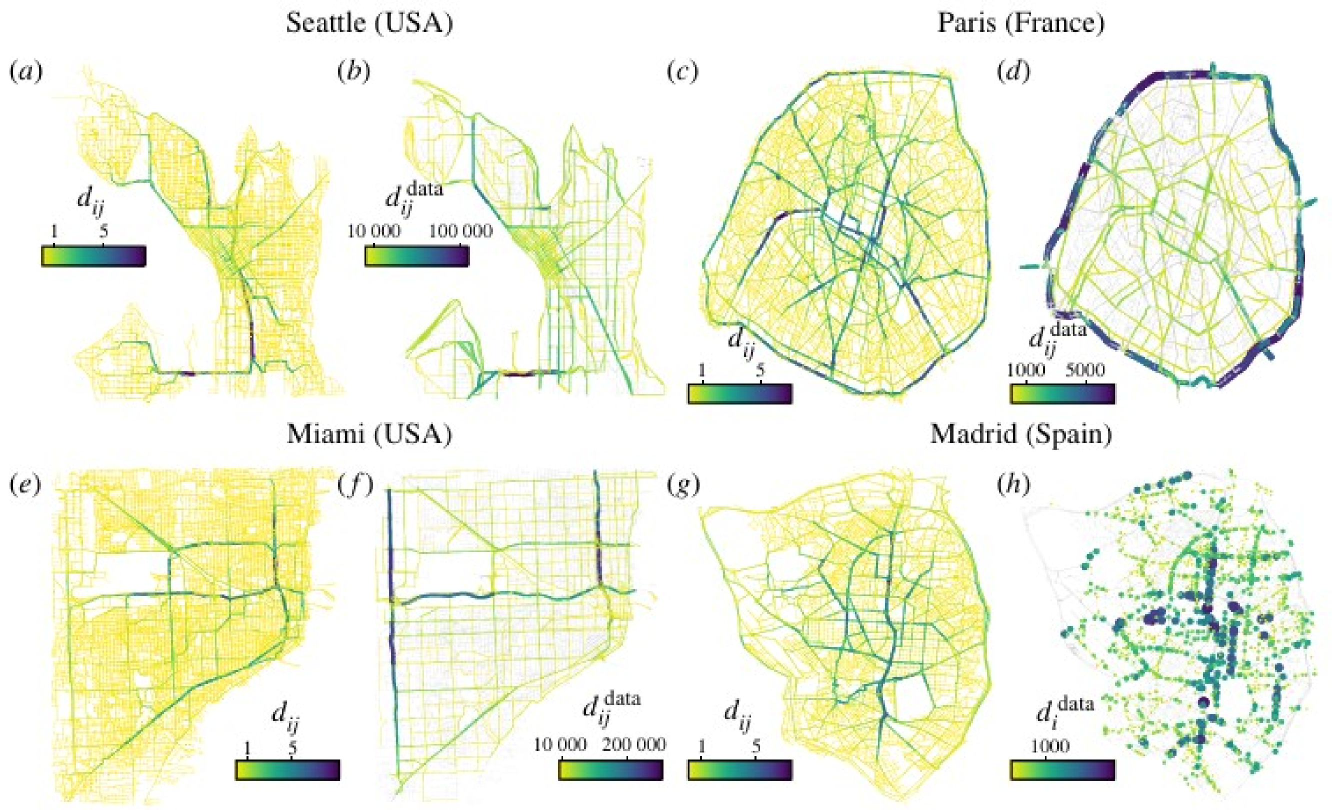
Congestion Hotspot Analysis

- Detection of congestion points and bottlenecks
- Impact of special events and weather conditions
- Practical implementations (e.g., Singapore, London)



Congestion Hotspot Analysis - Detection of Congestion Points and Bottlenecks

- Congestion hotspots are locations where traffic demand exceeds capacity, causing queues and delays.
- Methods of detection:
  - Loop detectors & radar sensors: measure speed and volume to identify recurring bottlenecks.
  - GPS/Floating Car Data (FCD): reveals travel time variability and localized slowdowns.
  - oMachine learning clustering techniques: automatically classify congested vs. non-congested links in real-time.
- Outcomes: hotspot mapping enables targeted infrastructure improvements (e.g., adding turning lanes, redesigning junctions).



A link model approach to identify congestion hotspots. Aleix Bassolas, Sergio Gómez, Alex Arenas. https://royalsocietypublishing.org/doi/10.1098/rsos.220894

Congestion Hotspot Analysis - Impact of Special Events and Weather Conditions

- Special events (sports, concerts, parades): generate non-recurrent congestion with atypical temporal/spatial demand.
- Weather impacts:
  - Rain, snow, and fog reduce roadway capacity, increase headways, and lower speeds.
  - ML models (e.g., Bayesian Networks, LSTM) can distinguish these anomalies from normal peak periods.
- Real-time hotspot analysis allows authorities to implement temporary control strategies such as contraflow lanes, variable message signs (VMS), and rerouting recommendations.

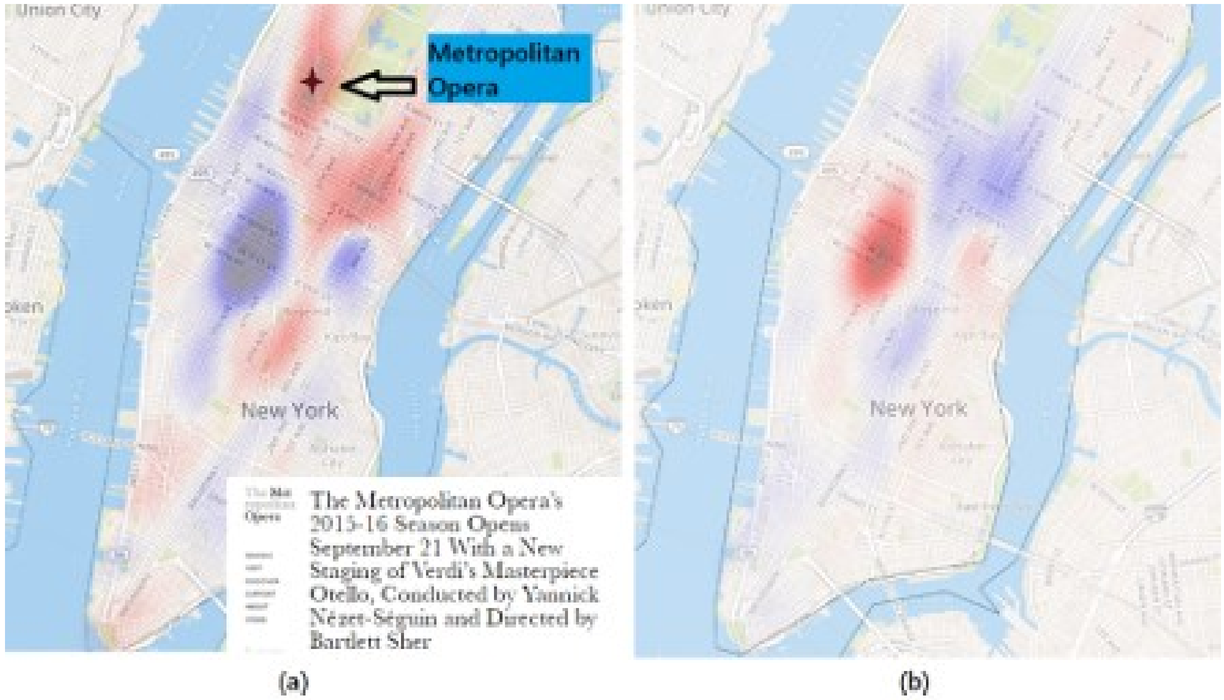
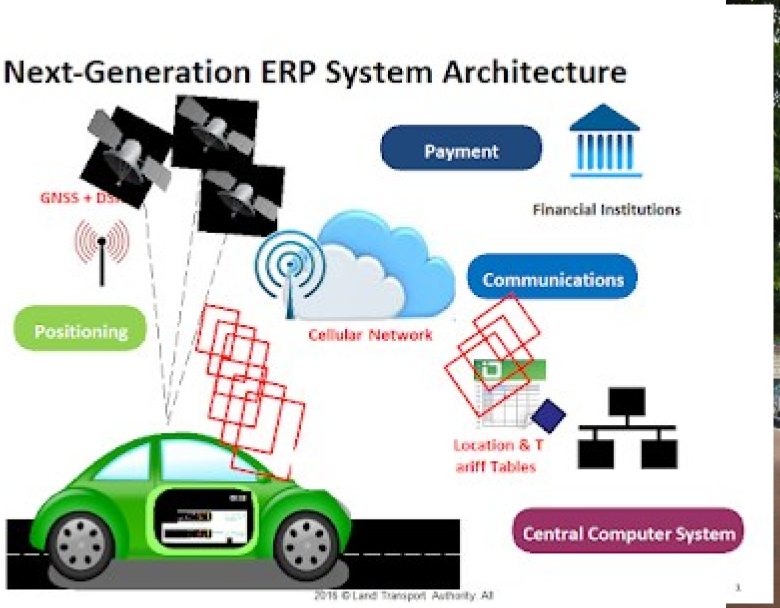


FIGURE 2.7: Differences in demand on (a) September 21 and (b) September 7.

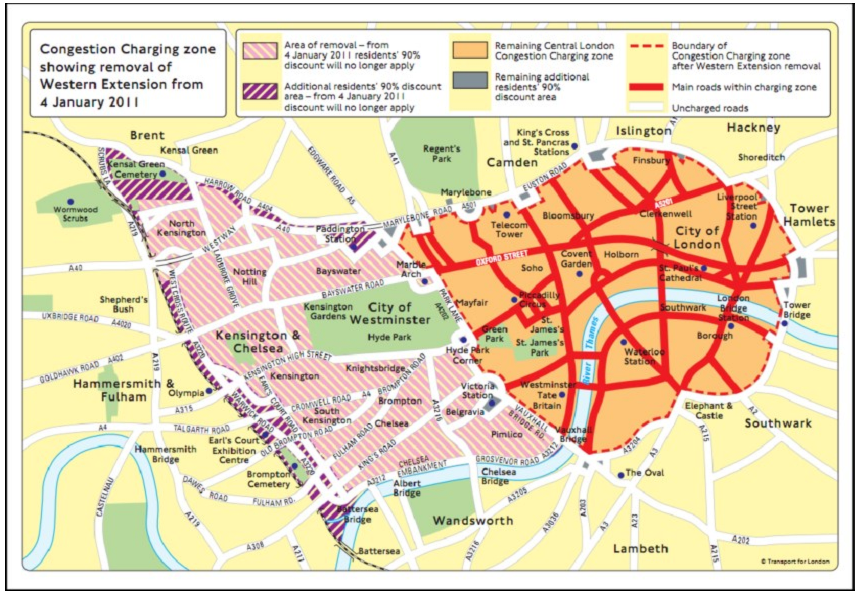
Figure: Taxi trajectory density map showing congestion hotspot during non-recurrent event (Pope's visit at Saint Patrick's Cathedral). Source: Markou (2020), "Detection, analysis and prediction of traffic anomalies due to special events"

Congestion Hotspot Analysis - Practical Implementations

- Singapore:
  - Extensive deployment of ERP (Electronic Road Pricing) and real-time monitoring systems.
  - Hotspot maps used for congestion charging and adaptive tolling, leading to 20–25% reduction in peak congestion.
- London:
  - Congestion Charging Zone (CCZ) combined with real-time traffic monitoring.
  - Use of GPS and ANPR (Automatic Number Plate Recognition) data to identify and monitor traffic hotspots.



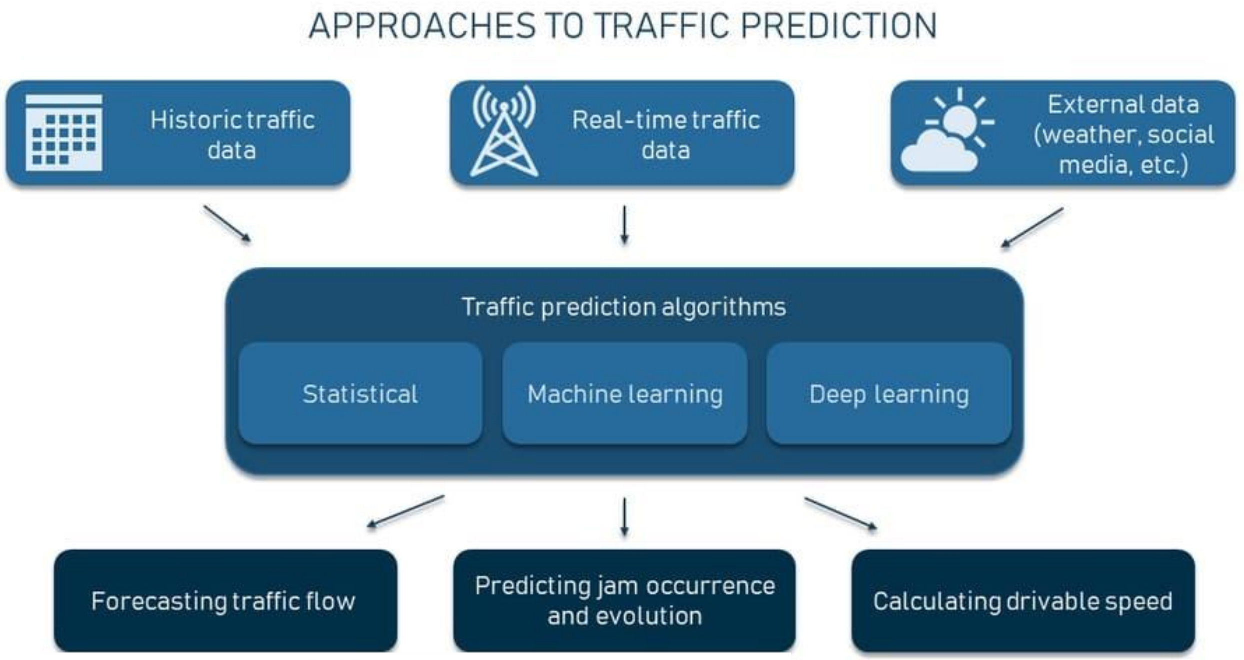
https://roadpricing.blogspot.com/2016/03/singapore-will-have-worlds-first-gnss.html



Tang, Cheng. (2020). The Cost of Traffic: Evidence from the London Congestion Charge. Journal of Urban Economics. 121. 103302. 10.1016/j.jue.2020.103302
.

Conclusions & Future Perspectives - Key Takeaways

- Enhanced precision: Machine Learning (ML) significantly improves the accuracy of traffic flow and travel time predictions compared to traditional statistical models.
- Data-driven insights: Integration of diverse data sources (loop detectors, GPS, cameras, social media) enables detection of recurrent patterns and non-recurrent anomalies.
- Operational benefits: Real-time prediction supports adaptive signal control, congestion hotspot management, and more reliable travel times, often reducing delays by 10–20% in practice.



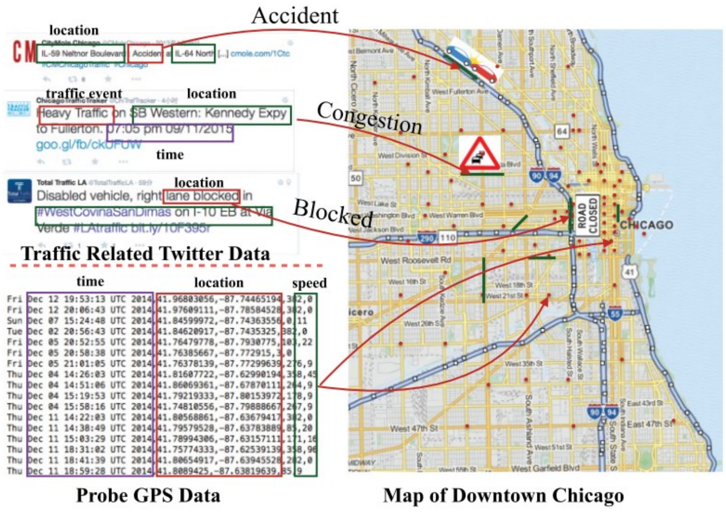
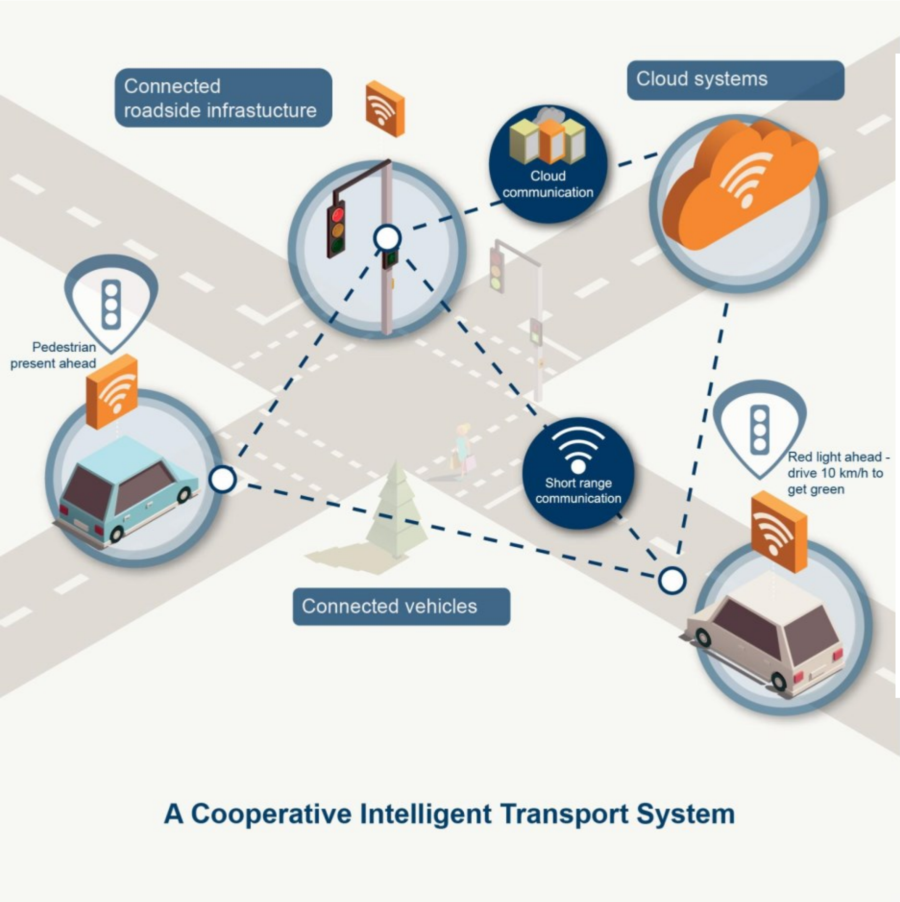


Fig. 1. Illustration of probe GPS data and social media data for traffic monitoring.

https://www.altexsoft.com/blog/traffic-prediction

Conclusions & Future Perspectives - Future Outlook

- Integration with Cooperative Intelligent Transport Systems (C-ITS):
  - Real-time exchange of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) data.
  - Improved coordination across multiple agencies and jurisdictions.
- Autonomous Vehicles as Mobile Sensors:
  - Large-scale data streams from connected autonomous vehicles (CAVs).
  - Potential for high-resolution, low-latency traffic monitoring.
- Edge & Cloud Computing:
  - Decentralized, scalable processing of massive data flows.
  - Supports immediate adaptive responses at the network level.



https://www.qld.gov.au/transport/projects/cavi/connected-vehicles

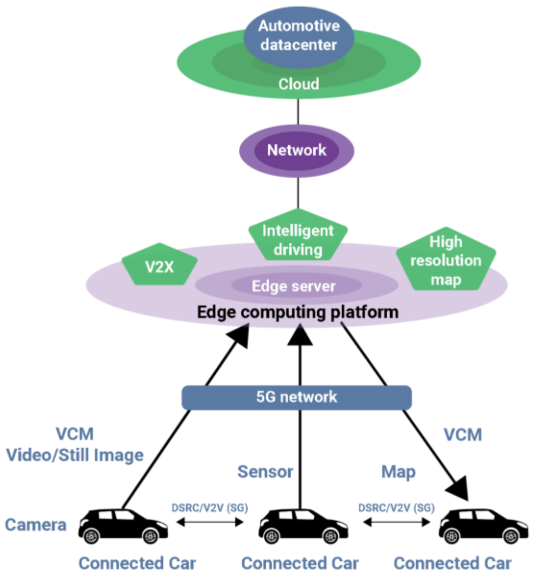


Figure 5 - Collabprative edge-cloud computing

https://www.gsaglobal.org/forums/edge-ai-computing-advancements-driving-autonomous-vehicle-potential

Conclusions & Future Perspectives - Open Questions for Discussion

- How can privacy-preserving techniques (e.g., differential privacy, federated learning) be effectively integrated into large-scale traffic forecasting?
- What governance and data-sharing models are needed for international C-ITS interoperability?
- How can we balance the explainability of ML models with the need for predictive accuracy in safety-critical traffic management?
- What role should public–private partnerships (e.g., with mobility service providers like Uber or Waze) play in shaping future data-driven mobility systems?