VEHICULAR 2026

The Fifteenth International Conference on Advances in Vehicular Systems, Technologies and Applications

ISBN: 978-1-68558-348-4

March 8th –12th, 2026

Valencia, Spain

VEHICULAR 2026 Editors

Dirceu Cavendish, Kyushu Institute of Technology, Japan

VEHICULAR 2026

Forward

The Fifteenth International Conference on Advances in Vehicular Systems, Technologies and Applications (VEHICULAR 2026), held between March 8-th, 2026 and March 12-th, 2026 in Valencia, Spain, continued a series of international events on considering state-of-the-art technologies for information dissemination in vehicle-to-vehicle and vehicle-to-infrastructure and focusing on advances in vehicular systems, technologies, and applications.

Mobility brought new dimensions to communication and networking systems, making possible new applications and services in vehicular systems. Wireless networking and communication between vehicles and infrastructure have specific characteristics from other conventional wireless networking systems and applications (rapidly changing topology, specific road direction of vehicle movements, etc.). These led to specific constraints and optimizations techniques; for example, power efficiency is not as important for vehicle communications as it is for traditional ad hoc networking. Additionally, vehicle applications demand strict communications performance requirements that are not present in conventional wireless networks. Services can range from time-critical safety services, traffic management, to infotainment and local advertising services. They introduce critical and subliminal information. Subliminally delivered information, unobtrusive techniques for driver's state detection, and mitigation or regulation interfaces enlarge the spectrum of challenges in vehicular systems.

We take here the opportunity to warmly thank all the members of the VEHICULAR 2026 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to VEHICULAR 2026. We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the VEHICULAR 2026 organizing committee for their help in handling the logistics of this event.

We hope that VEHICULAR 2026 was a successful international forum for the exchange of ideas and results between academia and industry for the promotion of progress in the field of vehicular systems, technologies, and applications.

VEHICULAR 2026 Chairs

VEHICULAR 2026 Steering Committee
Khalil El-Khatib, University of Ontario Institute of Technology – Oshawa, Canada
Éric Renault, ESIEE Paris, France
Xiaohong Peng, Birmingham City University, UK
Yuping He, University of Ontario Institute of Technology, Canada

VEHICULAR 2026 Publicity Chairs
Francisco Javier Díaz Blasco, Universitat Politècnica de València, Spain
Ali Ahmad, Universitat Politècnica de València, Spain

José Miguel Jiménez, Universitat Politècnica de València, Spain
Sandra Viciano Tudela, Universitat Politècnica de València, Spain

VEHICULAR 2026
Committee

VEHICULAR 2026 Steering Committee

Khalil El-Khatib, University of Ontario Institute of Technology – Oshawa, Canada
Éric Renault, ESIEE Paris, France
Xiaohong Peng, Birmingham City University, UK
Yuping He, University of Ontario Institute of Technology, Canada

VEHICULAR 2026 Publicity Chairs

Francisco Javier Díaz Blasco, Universitat Politècnica de València, Spain
Ali Ahmad, Universitat Politècnica de València, Spain
José Miguel Jiménez, Universitat Politècnica de València, Spain
Sandra Viciano Tudela, Universitat Politècnica de València, Spain

VEHICULAR 2026 Technical Program Committee

Vahdat Abdelzad, University of Waterloo, Canada
Nor Fadzilah Abdullah, National University of Malaysia (UKM), Malaysia
Mohammed Al-Ansi, University Malaysia Perlis (UniMAP), Malaysia
Sufyan T. Faraj Al-Janabi, University of Anbar, Ramadi, Iraq
Ali Alfoudi, Al-Qadisiyah University, Irak
Ala'a Al-Momani, Ulm University, Germany
Adel Aneiba, Birmingham City University, UK
Waqas Aman, Sultan Qaboos University, Oman
Muhammad Asim, Chung-Ang University, Seoul, South Korea
Hakan Aydin, Karadeniz Technical University, Turkey
Manlio Bacco, CNR-ISTI, Italy
Andrea Baiocchi, University of Roma "Sapienza", Italy
Paolo Barsocchi, ISTI (Institute of Information Science and Technologies) | Italian National Research Council (C.N.R.), Pisa, Italy
Paulo C. Bartolomeu, University of Aveiro, Portugal
Marcel Baunach, Graz University of Technology, Austria
Rahim (Ray) Benekohal, University of Illinois at Urbana-Champaign, USA
Sylvia Bhattacharya, Kennesaw State University, USA
Neila Bhouri, Université Gustave Eiffel, France
Konstantinos Blekas, University of Ioannina, Greece
Eugen Borcoci, National University of Science and Technology POLITEHNICA Bucharest, Romania
Alexandros-Apostolos A. Boulogeorgos, University of Piraeus, Greece
Christos Bouras, University of Patras, Greece
Alyssa Byrnes, University of North Carolina at Chapel Hill, USA
Roberto Caldelli, CNIT, Florence, Italy
Rodrigo Capobianco Guido, São Paulo State University (UNESP), Brazil
Antonio Carcaterra, Sapienza University of Rome, Italy

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

RoadSense – An AI-Based Vehicle Alert System Driven by V2V Mesh Communications

Amit Resh, Yigal Hoffner, Dean Itzhak, Tomer Ifargan, Shir Amar
Dept. of Computer Engineering
Shenkar - Engineering. Design. Art.
Ramat-Gan, Israel
e-mail: amit.resh@shenkar.ac.il, yigal.hoffner@shenkar.ac.il, deanitzhak@gmail.com, tomer.ifargan@stu.shenkar.ac.il, shiramar0401@gmail.com

***Abstract*— Vehicle-to-Vehicle (V2V) communication has long been recognized as a key player in cooperative safety, extending a vehicle's hazard awareness beyond the limits of local sensors. However, most prior studies address only one aspect of this challenge, either optimizing communication performance (latency, reliability, congestion) or training decision-making agents. This paper presents *RoadSense*, an integrated V2V hazard detection and response system designed to improve driving safety by extending awareness beyond local sensors using a mesh communication network and an AI-based agent. The AI agent is trained with reinforcement learning on a hybrid dataset of real V2V logs and domain-randomized simulations to predict and respond to potential forward collisions by issuing graded braking warnings.**

Keywords - Vehicle-to-Vehicle (V2V) Communication; Cooperative Collision Avoidance; Multi-agent reinforcement learning (MARL); Hazard Detection; Multi-Hop Mesh Network.

I. INTRODUCTION

Traffic congestion ahead of a vehicle can develop suddenly, leading to rapid deceleration or abrupt stops. Chain-reaction collisions frequently occur when a leading vehicle brakes unexpectedly, and following drivers have insufficient time or context to respond. While the first trailing car may detect and react, those farther behind often respond too late, typically only after the vehicle immediately ahead slows or collides.

A driver's or local controller's perception is inherently limited to what lies within direct view. This limitation worsens in adverse conditions, such as fog, rain, curves, or visual obstructions from trucks or buildings. Modern sensors, cameras, radar, ultrasonic units, and Light Detection & Ranging (LiDAR) extend local awareness but remain bound by physical range and line-of-sight constraints. Consequently, hazards hidden by terrain, traffic, or weather can remain undetected [1].

Vehicle-to-Vehicle (V2V) communication offers a complementary means of overcoming these limitations. By sharing safety-critical motion and hazard data among vehicles, V2V systems extend situational awareness beyond visual perception. Early hazard notifications increase available reaction time for both drivers and automated controllers, thereby reducing collision risk.

As shown in Figure 1, the control loop is extended by integrating information from an ad hoc vehicle-based mesh network. This addition transforms the process from reactive to predictive, enabling the ego-car to anticipate conditions beyond the range of its local sensors.

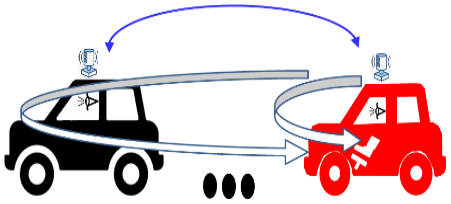

Figure 1. Ego-car extended driver-control loop with V2V mesh communication: data from preceding traffic augments local perception, extending the driver's awareness.

A. The RoadSense Project Goal

This project proposes an integrated hardware-software system that extends a vehicle-based mesh network and an AI-driven decision agent to process real-time data from the host vehicle and surrounding vehicles, including those beyond line of sight, providing timely alerts to the driver to apply brakes and prevent potential forward collisions (Figure 1). By continuously analyzing these inputs, the decision-making process can evaluate evolving traffic conditions and issue real-time alerts regarding potential hazards ahead. This approach enhances driving safety by enabling early detection of potential hazards over extended distances, thereby allowing braking intervention before hazards become critical.

The expected safety improvements are:

- Early alerts: Vehicles share warnings about hazards before drivers or local sensors detect them.
- Faster response: Extra time to react promptly lowers crash risk.
- Extended visibility: Hazards can be detected beyond curves, trucks, hills or other obstructions.

B. The RoadSense Research and Development Approach

Although the *RoadSense* system encompasses all aspects of an in-vehicle hazard detection and response system, the project's focus lies in four interrelated domains:

- Dynamic V2V communication: Developing a flexible ad-hoc wireless mesh that connects moving vehicles without relying on fixed infrastructure, enabling continuous exchange of motion data.
- Large-scale traffic scenario simulation: Generating diverse, controlled driving scenarios - both real and

synthetic - to train and test the decision-making process under variable conditions.
- AI-driven decision-making: Employing Reinforcement Learning (RL) to train an AI agent that evaluates dynamic traffic states and issues predictive braking or warning signals based on evolving conditions.
- Real-time integration and validation: Deploying and evaluating the trained system in real driving environments to assess its capacity for timely, reliable hazard detection and collision prevention.

C. The RoadSense Contribution

RoadSense advances the state of V2V hazard detection by implementing a complete closed loop, from multi-hop communication through learned hazard assessment to physics-based ego-car response, rather than stopping at packet-level metrics or scripted triggers. It introduces a four-layer realism framework that combines calibrated sensor noise, mesh-network impairments, environmental factors, and ego-car dynamics. This is supported by a hybrid dataset that combines real V2V logs with SMARTS (Scalable Multi-Agent Reinforcement Learning Training School for Autonomous Driving) based, domain-randomized simulations. The resulting system learns graded braking and warning policies using reinforcement learning and demonstrates low-cost deployability on a reproducible ESP32 (Espressif 32-bit Microcontroller) + GPS (Global Positioning System) + IMU (Inertial Measurement Unit) hardware platform, thereby linking realistic simulation to practical implementation.

D. The Structure of the Article

While *RoadSense* ultimately aims for a fully operational in-vehicle system, this paper emphasizes the development phase, during which the AI-based process is trained to recognize hazards and determine appropriate responses.

The article is structured as follows. It begins by describing the motivation for V2V-based hazard detection and reviews prior communication-centric and decision-centric research, establishing the gaps that *RoadSense* addresses. The system architecture is then presented, covering the mesh communication layer, the data-processing and filtering pipeline, the state-vector formulation, and the AI-based decision and action modules. The methodology section describes the development process, including the integration of prior knowledge from real V2V recordings, large-scale scenario simulation, and training of a reinforcement learning model, followed by a comparative analysis demonstrating improvements over existing approaches. The implementation and evaluation results are then presented for both simulated and real-world conditions. The paper concludes with a summary of contributions and potential future extensions.

II. Related Work and Comparative Context

Early research on V2V communication for driving safety focused primarily on demonstrating the feasibility and benefits of inter-vehicle data exchange for collision avoidance.

Yang et al. [2] introduced one of the first congestion-aware V2V protocols for Cooperative Collision Warning (CCW), demonstrating that low-latency, differentiated message delivery could significantly reduce reaction times during emergency braking. Later systems, such as COVCRAV [3], extended this concept to cooperative hazard signaling, introducing user-reported road hazards via interactive driver interfaces. Other simulation-based studies, such as Xie et al. [4], have demonstrated that V2V-augmented braking systems can outperform radar- or camera-based Automatic Emergency Braking (AEB) systems, particularly in limited-visibility conditions. Similarly, Joerer et al. [5], used the Veins simulator to show that inter-vehicle communication substantially reduces intersection collisions compared to perception-based control.

While these works established the communication benefits of V2V safety systems, they primarily relied on scripted hazard events and predefined control logic, rather than learning from contextual driving data. Over the past two decades, vehicle-safety research has advanced along two main directions: **communication-centric** and **decision-centric** approaches.

Communication-centric studies: typically based on the SUMO: Simulation of Urban MObility (SUMO) - Vehicles in Network Simulation (Veins) - Objective Modular Network Testbed in C++ (OMNeT++) (SUMO–Veins–OMNeT++) simulation stack, focus on optimizing network performance metrics such as Packet Reception Ratio (PRR), latency, jitter, channel load, and penetration rate. In these works, “hazards” are typically modeled as predefined events with scripted vehicle responses, thereby avoiding the deeper question of how a system should identify or evaluate hazards in context. Within this stack, SUMO models microscopic traffic mobility [6], [7]; OMNeT++ simulates the communication layer (using INET or Simu5G) [8]; and Veins synchronizes traffic and communication states [9], [10]. Together, they enable rigorous evaluation of V2V/V2X protocols such as IEEE 802.11p/ITS-G5 and C-V2X/NR-V2X under realistic traffic motion, leaving aside learned hazard assessment and vehicle-level control.

In contrast, **decision-centric studies** - most commonly using CARLA [11], [12] or SMARTS [13], [14] - train Reinforcement-Learning (RL) or autonomous-driving agents to make braking, yielding, or lane-change decisions in physics-based environments. However, these frameworks typically assume idealized communication conditions or omit V2V messaging altogether, leaving the learned policies unexposed to realistic network effects such as delay, loss, duplication, or out-of-order delivery.

More recently, **co-simulation efforts** have begun to bridge mobility dynamics and V2X networking—for example, MOSAIC with CARLA and SUMO with OMNeT++ setups [15], [16] coupled with physics engines or brake-control hardware. These studies achieve real-time, closed-loop integration in which communication signals influence vehicle dynamics. However, most still compromise along one axis: some employ learned decision-making under

simplified communications, while others model realistic V2V impairments but rely on hand-crafted hazard logic (e.g., rule-based triggers, Time-To-Collision (TTC) thresholds). Consequently, few studies train and evaluate learned hazard-assessment policies under entirely realistic conditions that simultaneously incorporate calibrated sensor noise, network impairments, environmental effects, and physics-accurate ego-car dynamics.

Despite significant progress in V2V communication research, most prior studies have focused either on network-level metrics (latency, packet delivery, or congestion) or on scripted decision logic. Few have addressed the whole pipeline from communication through learned hazard assessment to physically accurate vehicle responses. The *RoadSense* project directly addresses this gap and distinguishes itself from prior research on V2V safety systems by integrating multiple domains of communication, AI decision-making, and physical realism into a unified, functional system within a reproducible, low-cost platform.

RoadSense differs from prior work along the following issues:

- Closed loop (comms → decision → action): *RoadSense* extends beyond packet-level metrics and scripted hazard triggers by *learning* hazard assessment via RL and linking it directly to a physics-based ego-response model.
- Four-layer realism: Unified modeling of sensor noise, mesh-network impairments, environmental conditions, and ego dynamics, all calibrated using an ESP32 + GPS + IMU testbed - rarely achieved concurrently.
- Hybrid data + domain randomization: Combines real V2V logs with SMARTS-based simulations to narrow the sim-to-real gap, providing a stronger rationale than simulation-only or proprietary crash corpora.
- Learned hazard assessment: Reinforcement learning (Deep Q-Network (DQN) → Proximal Policy Optimization (PPO)) maps the ego-car's plus neighboring vehicle's state vectors to graded braking and warning actions integrated in the complete closed loop - a capability usually explored in isolation elsewhere.
- Graded action policy: Beyond binary alert/no-alert logic, the learned policy maps confidence to graded braking and driver visualization or actuation, enabling interpretable, adaptive responses.
- Low-cost deployability: Implements a reproducible hardware reference design (ESP32 microcontroller, NEO-6M GPS, Motion Processing Unit model MPU-6050) demonstrating real-world feasibility rather than remaining a purely software-stack concept.

III. System Usage in the Deployment Phase

The operational flow in the final system within the car, once communication processes and AI-based decision-making have been trained and deployed, is shown in Figure 2:

- Mesh communications: An ad-hoc V2V network that provides the vehicle with data from nearby cars, including those beyond line of sight (e.g., hidden by curves, trucks, or weather).
- Data collection and transformation: Gathers and merges data from surrounding vehicles with the ego-car's sensor data, removes duplicates and irrelevant information, and converts it into state vectors for an AI-based decision-making process.
- Decision-making: The trained AI decision process analyzes both local and received data to identify potential hazards ahead based on extensive reinforcement learning on a combination of real-world and simulated driving data.
- Action-taking: Based on the AI's assessment, this process determines the appropriate response - specifically, whether braking is necessary and, if so, the recommended braking intensity.
- Warning: Presents timely input to the braking mechanism or visual or auditory cues to the driver via the vehicle's dashboard interface.

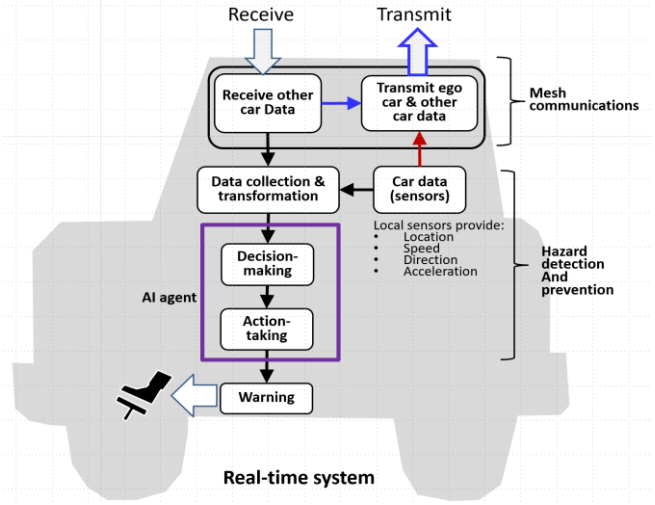

Figure 2. Architecture of the RoadSense real-time system: Sensor and V2V data are joined to support AI-based decision-making on braking commands.

A. Mesh (Multi-Hop) V2V Communication Framework

Each vehicle transmits its local sensor data to surrounding vehicles. To avoid communication congestion, as described by Xue Yang et al. [2], the transmission size was kept to a bare minimum, including only GPS location, acceleration, and identification. Nearby vehicles retransmit both their own data and the data they receive (and filter) to others in their vicinity, as shown in Figure 3. This creates an ad hoc, extensible mesh network in which messages are forwarded hop-by-hop across multiple vehicles. Such a multi-hop communication scheme significantly extends the effective range of data exchange, allowing each car to receive information from vehicles well beyond its direct line of sight and beyond the direct transmission distance.

For example, a car 1 km ahead can send data to the ego-car via intervening vehicles, even though it is beyond the

ego-car's direct range. The mesh's redundancy ensures that if one path fails, the message can still be delivered via alternative routes, improving reliability.

B. Data Collection and Transformation Process

Upon receiving incoming data, the system standardizes and validates raw inputs from surrounding vehicles, removing duplicate, irrelevant, and incomplete information to ensure high-quality data for decision-making. The data is then time-aligned and transformed into the ego-car frame. In addition, the system performs consistency checks on GPS coordinates, timestamps, and sensor reliability to ensure that only verified, accurate data are processed and used.

Figure 3. Mesh V2V communications: Vehicles broadcast data to nearby cars, which rebroadcast it with their own data, forming a multi-hop mesh that propagates messages.

1) Cone-based filtering

Before entering the AI decision-making process, all incoming data from the mesh network undergoes filtering to retain only data relevant to the AI predictions. Each vehicle transmits its acceleration and GPS location. For hazard detection, only vehicles traveling ahead in the same lane and direction within a specified distance are considered relevant, thereby defining a cone-shaped region of interest extending forward from the ego-car, as shown in

Figure 4. The AI component analyzes the motion vectors of nearby vehicles within this region to assess potential collision risks and congestion trends. By discarding irrelevant, redundant, duplicate, and distant data, the filtering process ensures that only contextually significant information is retained for trajectory analysis and decision-making, thereby allowing each vehicle to operate efficiently without overloading the prediction algorithm. In addition, to further reduce unnecessary communication congestion, only messages that passed the filtering are retransmitted.

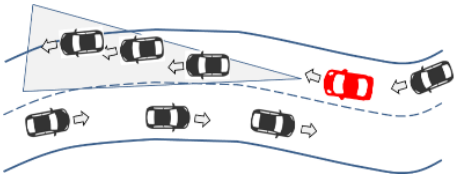

Figure 4. Transmitted car data filters out irrelevant cars by focusing on a cone-shaped area in the direction of travel.

2) State Vector Generation

Input into the decision-making process is provided by a state vector that represents the ego-car's state and the states of surrounding vehicles. The state vector includes:

- Ego-car parameters: speed, acceleration, azimuth, lane index/offset, and position in the ego-centric frame. All calculated locally from consecutive GPS locations and an accelerometer sensor.
- Traffic-context parameters (up to N neighbors): for each neighboring vehicle, local calculations generate relative distance and bearing, relative velocity and acceleration, with respect to the ego-car. This context encodes how the ego-car is positioned and moving relative to others; for example, whether a vehicle is rapidly decelerating sharply ahead, and how fast and how close the ego-car is to others ahead.
- Auxiliary/environmental parameters (optional): speed limit, curvature, road friction estimate, weather/visibility flags, given that they are available.

C. Decision-making Process

This is the AI-based process resulting from extensive reinforcement learning on real and simulated driving data, which will be explained in detail in the development phase section.

D. Action-taking Process

The *Action-Taking* process interprets the raw output of the decision-making process and converts it into human-readable or system-level commands. It does not generate its own predictions; instead, it maps the AI's continuous confidence values into discrete, safety-oriented actions and alert categories. Rule-based logic derived from kinematic safety models governs this mapping:

- 0.0–0.3 → No alert (safe distance maintained)
- 0.3–0.6 → Caution alert ("Light Brake Needed")
- 0.6–1.0 → Emergency alert ("Immediate Brake Required")

These thresholds can be dynamically adjusted based on contextual factors such as road friction, speed limits, predicted time-to-collision, and weather (Figure 5). The decision logic may also select the appropriate alert modality (visual, audio, or haptic) and verify that braking recommendations remain stable across consecutive frames before activation, ensuring both reliability and driver trust.

E. Warning Process

The *RoadSense* project defines warnings (or alerts) as the core output of the system's hazard-detection logic. Their role is to:

- Warn drivers of imminent risks such as potential collisions, sudden braking, or unsafe speeds.
- Provide graded or prioritized notifications, distinguishing critical, cautionary, and convenience-level messages.
- Improve reaction time by informing drivers before hazards are visible to sensors or the human eye.

The alerts are intended to enhance situational awareness rather than replace driver control; the system serves as a driver-assistance layer rather than an autonomous intervention mechanism.

IV. AI TRAINING IN THE DEVELOPMENT PHASE

The ego-car's decision-making process relies on an AI model trained to accurately predict brake-warning signals by analyzing both local sensor inputs and remote data streams received from upstream vehicles via a mesh network. To effectively train this model, a simulation environment is constructed to expose it to a wide range of driving scenarios, thereby enabling robust learning of collision-avoidance behaviors.

Training involves generating data for both the ego-car and the surrounding vehicles, as shown in Figure 5. Moreover, the ego-car's physical dynamics must be simulated in real time to generate data and simulate the car's reaction whenever the decision-making process initiates a braking action.

The primary focus of the training phase is on the following key processes shown in Figure 5:

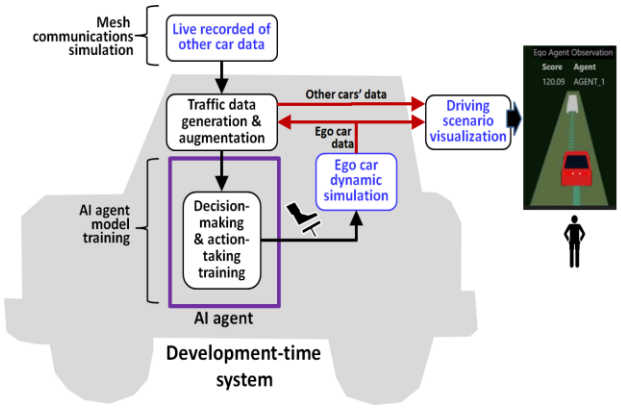

Figure 5. In the *RoadSense* simulation framework, recorded and synthetic data are used to train the decision-making AI.

- **Traffic Scenarios Data Recording and Augmentation**: Actual message communications are recorded in vehicles deployed with sensors and used to generate everyday traffic scenarios for simulation. Because real crash data are rare, simulated crashes and variations of traffic scenarios are generated to enrich the simulator dataset.
- **AI Agent Model Training**: The real and simulated driving-generated data is fed into the AI model that is trained using reinforcement learning to detect and respond to hazards.
- Ego-car Dynamics Simulation: In the simulator, a physical model of the ego-car and its dynamic properties is employed to replicate real driving behavior and calculate the vehicle's physical response to the braking actions recommended by the decision-making process.
- The simulation serves two primary purposes:
  - **Physics-Based Scenario Replay**: It integrates simulated data into the data collection pipeline, providing accurate representations of the ego-car's physical state. This enables realistic progression through driving scenarios by combining recorded or generated data from surrounding vehicles with physics-based modeling of the ego-car's behavior during critical phases, such as braking.
  - **AI Model Evaluation and Training**: It facilitates assessment of the decision-making process by determining whether the AI avoids collisions and maintains safe driving conditions. This feedback loop supports reinforcement learning, allowing the AI model to improve its control strategy based on the simulated outcomes of each decision step.
- **Driving Scenario Visualization**: During development, a visualization module was implemented to display real-time vehicle positions and trajectories following braking or warning actions initiated by the decision-making system. This visualization enables assessment of driving outcomes and potential collision scenarios, providing valuable feedback on the system's effectiveness and response accuracy.

The above processes and their design are described in greater detail in the following sub-sections.

A. The Simulation Environment and Data Generation

The simulator is designed to utilize data that closely reflects real-world driving conditions. Vehicles equipped with dedicated hardware record GPS coordinates, accelerometer readings, and timestamps while operating on actual roads. These recordings capture diverse scenarios, including approaching intersections, navigating curves, and driving in low-visibility conditions. Additionally, the vehicles implement mesh communication protocols and log transmissions received from nearby cars. This collected data serves as the foundation for generating realistic simulation scenarios that include trajectories and mesh communications as expected in the real world.

In addition to collecting real-world data, synthetic data is programmatically generated to support various testing and training objectives. This includes increasing traffic density through data replication, simulating sudden braking events and full-stop collisions, and introducing communication anomalies such as increased latency, message loss, and network congestion.

The ego-car's motion is simulated using its physical parameters and kinematic equations to model its trajectory before and after any braking command issued by the decision-making module. Initial path and velocity vectors are specified as input parameters and are dynamically updated based on the AI-driven decision process's evolving outputs.

Traffic scenarios are implemented in SMARTS, an open-source simulator for multi-agent reinforcement learning [13] extended with our own communication and noise-modeling modules. Instead of relying on proprietary accident datasets, real-world constraints are embedded directly into the simulation. The scenario scripts:

- Select a road segment (urban arterial or multi-lane highway) and legal-speed geometry.
- Sample traffic density and vehicle type distribution.

- Initialize actors with lane, speed, headway, and position relative to the ego-car.
- Trigger events (e.g., hard breaking, cut-ins, stop-and-go waves, obscured obstacles) in randomized time windows.

To promote generalization, we apply domain randomization within realistic bounds informed by published driving statistics and our hardware measurements. Parameters varied include vehicle speeds (±10–20%), packet loss (0–30%), message-delay jitter, GPS noise at the meter scale, and standard lane/vehicle dimensions.

The resulting environment exposes the AI agent to diverse hazard scenarios, such as rapid deceleration, high-density traffic, and occluded obstacles. It enables safe learning from situations that cannot be recreated in real vehicles. Simulated and real V2V data from our testbed are combined to form a comprehensive training set, thereby improving accuracy and robustness prior to deployment.

B. *Deep Q-Learning and PPO-Based Adaptive Decision-Making*

We implemented a Reinforcement Learning (RL) model in *RoadSense* to issue timely braking recommendations to avoid collisions.

The RL objective rewards correct hazard responses (timely slow/stop to avoid collisions) and penalizes failures (late response, collisions, unnecessary harsh braking).

Two reinforcement learning algorithms were employed:

1) Deep Q-Learning (DQN): used during early development to train the agent on discrete binary actions (e.g., brake / no brake).

2) Proximal Policy Optimization (PPO): adopted for the final system due to its superior performance in continuous control domains and its robustness against unstable policy updates, Schulman [17], Minih [15]. PPO constrains policy updates within a trust region, ensuring stability and preventing overfitting to transient simulation states. Simulation runs performing RL training for interactions between 3 cars: an ego-car and two cars in front, have been shown to successfully converge after several hundred-thousand Epochs, as shown in Figure 6.

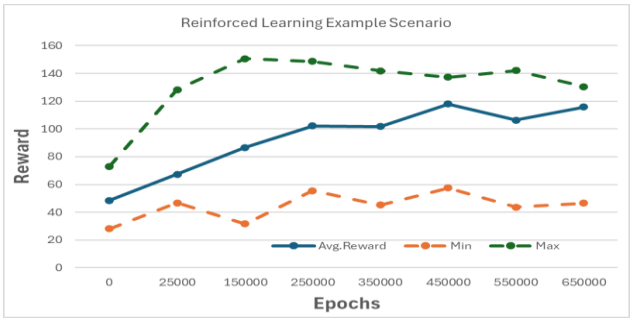

Figure 6. Example of a Reinforcement-Learning session that converged after 650K Epochs.

C. *Driving Scenario Visualization and Warning*

For visualization, a Unity-based environment is used to display vehicle positions and trajectories before and after braking actions, illustrating the vehicle's dynamic response to the decision-making system's braking or warning commands, as shown in

Figure 7. This enables the assessment of driving outcomes and potential collision scenarios, providing valuable feedback on the effectiveness and timing of the decision-making process.

V. ROADSENSE DEVELOPMENT, INSTALLATION AND TESTING

The initial phase involved assembling and validating multiple hardware boards, which were subsequently installed in vehicles and powered via the onboard USB interface. As part of the Mesh communications system, custom firmware was developed for the microcontroller to perform sensor monitoring, transmit telemetry data at 1 Hz, and operate as a node in an ad hoc mesh network. To reduce communication overhead, filtering mechanisms were implemented to selectively retransmit only data relevant to trailing vehicles, specifically, information originating from cars within the forward extended cone, as described earlier. Additionally, for development and debugging, the firmware was enhanced to monitor and log sensor readings and network communications.

The second phase involved capturing and recording real-world vehicular telemetry data across various driving scenarios, including deceleration before intersections and sharp turns. These raw recordings were subsequently processed and programmatically augmented to synthetically expand the dataset. This augmentation introduced variability and increased the diversity of traffic scenarios, resulting in a comprehensive database encompassing both typical driving patterns and artificially generated incident conditions.

In the third phase, the SMARTS simulation framework [13] was employed to emulate traffic conditions and mesh communications using the previously recorded dataset. Within the simulator, an ego-car was dynamically controlled using physical parameters to emulate realistic physical driving behavior. An integrated AI module was tasked with issuing predictive braking alerts, enabling the ego-car to decelerate smoothly and avoid collisions with vehicles ahead under varying traffic conditions. The simulator iteratively executed each traffic scenario over a substantial number of epochs, during which the AI module was trained using reinforcement learning techniques to optimize its decision-making policy.

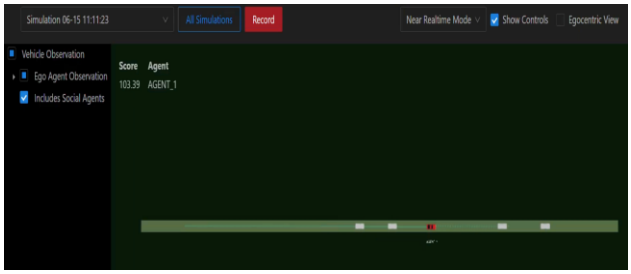

Figure 7. Visualization of the response of the ego-car to the warnings issued by the decision-making processes.

Initial validation of the trained model was conducted using an external visualization process built on the Unity engine. After the learning algorithm converged, this component was used to render graphical representations of each scenario's outcome, enabling qualitative assessment of the AI module's performance.

The final phase involved deploying the trained AI agent onto the ego-car's onboard hardware platform and conducting field tests under real-world driving conditions. Validation procedures focused on verifying the timing and accuracy of the brake alert relative to the human driver's decisions. The results demonstrated that, even when the driver's visibility of preceding vehicles was obstructed, the AI agent reliably initiated braking promptly, effectively mitigating collision risk.

VI. Results and Discussion

The *RoadSense* system was implemented and validated through a staged process combining simulation and road testing. During the simulation phase, the Reinforcement-Learning (RL) agent was trained using a hybrid dataset comprising both real V2V mesh recordings and SMARTS-generated traffic scenarios. Domain randomization was applied to communication latency, packet loss, GPS jitter, and environmental factors. The model successfully converged within approximately 650,000 training rounds, yielding stable hazard detection and braking policies.

During controlled simulations, the trained AI consistently generated timely, graded alerts across diverse driving contexts—including sudden braking, stop-and-go waves, and reduced-visibility conditions, demonstrating accurate state-vector interpretation and effective mapping of policy confidence to braking intensity.

Field deployment using a low-cost ESP32 paired with GPS and IMU sensors confirmed that the end-to-end mesh–decision–action loop functioned in real time, achieving end-to-end latency below 2 seconds. Tests under partial occlusion (e.g., blocked line of sight by larger vehicles) showed that the AI module could initiate braking before the driver's visual recognition of hazards, confirming predictive behavior consistent with simulation results. The graded alert mechanism improved driver reaction timing and system interpretability without false or unnecessary interventions.

VII. Summary, Conclusions and Future Work

In summary, existing research either optimizes the communication layer in isolation or develops decision-making agents under idealized conditions, rarely integrating both within a realistic, sensor-calibrated, physics-accurate framework. *RoadSense* presents a comprehensive, closed-loop V2V hazard detection system that integrates realistic communication models, physics-driven vehicle dynamics, and reinforcement-learning-based decision-making algorithms within an accessible, cost-effective platform.

Unlike previous studies that focus on either network metrics or scripted control logic, *RoadSense* merges four calibrated realism layers—sensor noise, mesh-network impairments, environmental variation, and vehicle dynamics into a single end-to-end architecture that learns hazard semantics rather than applying fixed thresholds.

The hybrid training methodology, which combines real-world mesh-network data with domain-randomized SMARTS simulations, enabled the RL agent to adapt to noisy communications and unpredictable driving conditions. Results confirm that learned policies can operate effectively on embedded hardware, providing graded alerts that enhance situational awareness and reduce driver reaction times even beyond line of sight.

Subsequent development will expand *RoadSense* to encompass additional sensors, more sophisticated mesh communication protocols, multi-vehicle coordination, and traffic scenarios that extend to additional hazardous conditions.

References

[1] F. d. P. Müller, "Survey on ranging sensors and cooperative techniques for relative positioning of vehicles," Sensors, vol. 17, p. 271, 2017.

[2] X. Yang, L. Liu, N. H. Vaidya and F. Zhao, "A Vehicle-to-Vehicle Communication Protocol for Cooperative Collision Warning," in 1st Annual International Conference on Mobile and Ubiquitous Systems (MobiQuitous 2004), Networking and Services, Cambridge, MA, USA, 2004.

[3] F. Outay, H. Bargaoui, A. Chemek, F. Kamoun and A Yasar, "The COVCRAV project: Architecture and design of a cooperative V2V crash avoidance system," in The 3rd International Workshop on Connected & Intelligent Mobility, Coimbra, Portugal, 2019.

[4] Z. Xie, D. Wu, D. Liu, H. Huang and B. Xie, "V2V Communication-based AEB Validation in Traffic Accident Simulation Scenario," in ICFEICT 2021: International Conference on Frontiers of Electronics, Information and Computation Technologies, Changsha, China, 2021.

[5] S. Joerer, B. Bloessl, M. Huber, A. Jamalipour and F. Dressler, "Assessing the impact of inter-vehicle communication protocols on road traffic safety," in Proceedings of the 6th annual workshop on Wireless of the students, by the students, for the students, Maui, Hawaii, USA, simmer.

[6] M. Behrisch, L. Bieker-Walz, J. Erdmann and D. Krajzewicz, "SUMO – Simulation of Urban MObility: An Overview," in SIMUL 2011, 2011.

[7] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent Development and Applications of SUMO—Simulation of Urban MObility," International Journal on Advances in Systems and Measurements, vol. 5, no. 3, pp. 128-138, 2012.

[8] R. H. A. Varga, "An Overview of the OMNeT++ Simulation Environment," in Proc. of the 1st International Conference on Simulation Tools and Techniques for Communications, Networks and Systems (SimuTools), 2008.

[9] C. Sommer, "Bidirectionally Coupled Network and Road Traffic Simulation for Improved IVC Analysis," IEEE Transactions on Mobile Computing, vol. 10, no. 1, pp. 3-15, 2011.

[10] C. Sommer et al., "The open source vehicular network simulation framework," in Recent Advances in Network Simulation, Springer, 2019, pp. 215-252.

[11] D. Alexey, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An Open Urban Driving Simulator," in Conference on Robot Learning (CoRL), 2017.

[12] R. Gutiérrez-Moreno, R. Barea, E. López-Guillén, J. Araluce and L. Bergasa, "Reinforcement Learning-Based Autonomous Driving at Intersections in CARLA Simulator," Sensors, 2022.

[13] M. Zhou et al., "SMARTS: An Open-Source Scalable Multi-Agent RL Training School for Autonomous Driving," in Proceedings of the 2020 Conference on Robot Learning, 2021.

[14] A. Rasouli, "Driving SMARTS Competition at NeurIPS 2022: Insights and Outcome," in NeurIPS 2022 Competitions Track, 2022.

[15] D. Cui, Y. Shen, H. Yang, Z. Huang, K. Rush, P. Huang and P. Bujanovic, "Extensible Co-Simulation Framework for Supporting Cooperative Driving Automation Research," Transportation Research Record, Journal of the Transportation Research Board, pp. 1067-1079, 2022.

[16] A. Marchetta, A. Coppola, M. Cinque, M. Fiorentino and G. N. Bifulco, "An Eclipse MOSAIC-Based Hardware-in-Loop V2X Co-Simulation Framework for CCAM Services," in IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), Bilbao, Spain, 2023.

[17] S. Zeadally, J. Guerrero and J. Contreras, "A tutorial survey on Vehicle-to-Vehicle communications," Telecommunication Systems, vol. 73, no. 3, pp. 469-489, 2020.

[18] K. N. Toshiyuki Hagiya, "Acceptability evaluation of inter-driver interaction system via a driving agent using Vehicle-to-Vehicle communication," in The 11th Augmented Human International Conference, Winnipegm Manitobam Canada, 2020.

[19] S. Banerjee et al., "Link-Adaptation for Improved Quality-of-Service in V2V Communication using Reinforcement Learning," in The Second International Conference on AI-ML Systems, Bangalore, India, 2023.

[20] Y. Song, H. Chunxi, X. Weiyin and H. Dengbo, "HMI Design for Chain-Braking Event Based on V2V Communication," in AutomotiveUI '23 Adjunct: Adjunct Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ingolstadt, Germany, 2023.

[21] X. Zhang, J. Li, J. Zhou et al., "Vehicle-to-Everything Communication in Intelligent Connected Vehicles: A Survey and Taxonomy," Automotive Innovation, vol. 8, p. 13–45, 2025.

[22] S. Lu, "Modeling dynamics of traffic flow, information creation and spread through Vehicle-to-Vehicle communications: A kinetic approach," International Journal of Non-Linear Mechanics, vol. 175, 2025.

[23] J. B. Kenney, "Dedicated short-range communications (DSRC) standards in the United States," vol. 99, no. 7, p. 1162–1182, 2011.

[24] ETSI, "Access layer specification for Intelligent Transport Systems operating in the 5 GHz frequency band, ETSI EN 303 797 V2.1.1," 2019. [Online]. Available: https://standards.globalspec.com/std/10146015/en-302-571.

[25] S. I. T. Standard, " Dedicated Short Range Communications (DSRC) Message Set Dictionary," 2015. [Online]. Available: https://doi.org/10.4271/J2735_201509.

[26] G. Kahn, A. Villaflor, Vitchyr Pong, Pieter Abbeel, Sergey Levine, "Uncertainty-aware reinforcement learning for collision avoidance," arXiv preprint, arXiv:1702.01182, 2017.

[27] L. Cao et al., "Driving-policy adaptive safeguard for autonomous vehicles using reinforcement learning," arXiv preprint, arXiv:2012.01010, 2020.

[28] H. Chen, X. Cao, L. Guvenc and B. Aksun-Guvenc, "Deep-reinforcement-learning-based collision avoidance of autonomous driving system for vulnerable road user safety," Electronics, vol. 18, no. 10, pp. 1952, 2024.

[29] S.H. Ashwin and R. Naveen Raj, "Deep reinforcement learning for autonomous vehicles: lane keep and overtaking scenarios with collision avoidance," International Journal of Information Technology, vol. 15, no. 7, 2023.

[30] A. M. A. Behzadan, "Adversarial reinforcement learning framework for benchmarking collision avoidance mechanisms in autonomous vehicles," 2018. [Online]. Available: https://arxiv.org/abs/1806.01368.

[31] S. Zeadally et al., "A tutorial survey on Vehicle-to-Vehicle communications," Telecommunication Systems, vol. 73, no. 3, 2020.

[32] X. Yang, L. Liu, N. H. Vaidya and F. Zhao, "A Vehicle-to-Vehicle communication protocol for cooperative collision warning," in The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2004, pp. 114-123.

Evidence of Unmet Demand for Public Charging Points for Electric Vehicle Users in the United Kingdom

Dinesh Chacko
Institute of Electrical and Electronic Engineers
Maidenhead, United Kingdom
Email: dinesh.chacko@ieee.org

Monika M. Wahi
Research and Data Lab
New Delhi, India
Email: mwahi@RADL.online

***Abstract*—The widespread adoption of electric vehicles is necessary for the United Kingdom to achieve its net zero target by 2050, but barriers exist if the electric vehicle charging infrastructure is not well-established. This study aimed to characterise the distribution of public electric vehicle charging points as compared to the distribution of electric vehicles in areas across the Scottish region of the United Kingdom served by 2 energy networks, and investigate any evidence of unmet demand for public charging points. Publicly available data on charging points (n = 2,810) and charging sessions (n = 579,294) were obtained and analysed. The region studied contained so-called blackspot areas where 100 or more electric vehicles were owned, but there were less than 10 public charging points available (termed "high demand" areas). Charging sessions in these areas, when compared to sessions lasting less than 30 minutes, had over twice the odds of lasting 30 minutes to 1 hour, almost 3 times the odds of lasting between 1 and 2 hours, and over 5 times the odds of lasting 2 or more hours compared to charging sessions outside high demand areas. Also, compared to work day times, in high demand areas, charging sessions had over 3 times the odds of starting in evening hours, almost 3 times the odds of starting during sleep times, and about 2 and a half times the odds of starting in early morning, compared to other areas. In conclusion, the findings suggest that the number of public charging points available to serve electric vehicle owners in this area appears to be inadequate.**

Keywords*—*Electric vehicles; United Kingdom; charging point infrastructure; net zero; charging sessions; charging point demand; smart charging.

I. INTRODUCTION

As of 2021, the United Kingdom (UK) government set a target to reduce carbon dioxide emissions by 68% from 1990 levels by 2030, and to achieve net zero emissions by 2050 [1–3]. Consequently, the UK Climate Change Committee (CCC) established strategies to promote improved sustainable energy use to meet net zero targets, with a specific focus on promoting the adoption of Electric Vehicles (EVs) [1–4]. The widespread adoption of EVs is seen as a necessary condition to enable the UK to achieve net zero, as passenger cars have been documented as the largest source of road transport emissions [2][5]. To that end, in 2022, 17% of the new cars purchased in the UK were EVs, and one million EVs were registered in the UK in 2021, exceeding expectations [6][7].

However, the absence of a coordinated rollout of an EV public charging infrastructure that complements the enthusiastic uptake of EVs in the UK has been identified as a significant barrier to overall EV adoption in the UK [6–8]. Currently, charging facilities are not evenly spread across the UK, leaving areas with a low prevalence of public Charging Points (CPs) [7][9]. Although it is known that some regions in the UK lack public CP access, research into how well the public EV charging infrastructure is meeting the demand of EV users in the UK is lacking. Research addressing this point is necessary to inform UK leaders as to whether EV adoption is outpacing the development of the public EV charging infrastructure.

In October 2022, the UK had nearly 35,000 CPs, and over the first nine months of that year, more than 1,200 new public rapid chargers and over 5,000 new public standard chargers were installed [7]. Over the same period, there were 250,000 new EV registrations in the UK, which means there was on average 1 new public standard charger for every 50 new vehicles [7]. The UK government has termed some regions "blackspots" due to the lack of public CPs, forcing EV owners to drive long distances to charge their vehicles, which is inconvenient and offsets the savings anticipated from owning an EV compared to a car that uses fossil fuels [7][9]. As an example, 31% of the UK's public CPs are in London, compared to only 3% being in the North East region [7]. This analysis helps bring insight into whether the public CP infrastructure in a region of the UK is able to meet the demand of EVs users in that region. The results can provide guidance on how to further develop the public CP infrastructure.

Energy networks in the UK are organised into regional collections of primary energy stations and substations. The two networks that serve Scotland are Scottish Power Energy Networks (abbreviated SP Energy) and Scottish and Southern Energy Networks (abbreviated SSEN) [10][11]. In networks in the UK such as SP Energy and SSEN, Primary Substations (PSs) directly feed from high voltage networks, and serve to step down energy transmitted through these high voltage networks from 33 to 11 kilovolts (kV) in order to enable energy to be distributed to homes and businesses [12]. SP Energy serves the areas of central and southern Scotland, North Wales, Merseyside, Cheshire and North Shropshire [11]. SSEN serves the north of Scotland and central southern England, and claims 3.9 million homes and businesses as customers [10]. Our analysis focuses on PS service areas in Scotland served by SP Energy and SSEN, and the prevalence of both EVs and public CPs in those service areas.

The aim of this research was to 1) characterise the distribution of public CPs and EVs in PS service areas across the

regions served by SP Energy and SSEN, and 2) investigate if there is evidence of unmet demand for public CPs in PS service areas in these networks with a high prevalence of EVs coupled with a low density of public CPs. The structure of this paper proceeds with Section II, which describes our materials and methods, including the sources of the data we used and our analytic approach. In Section III, the results are presented, including a comparison of EV ownership and public CPs in PS service areas, an examination of densities of public CPs and EV ownership, and an analysis of charging session patterns in high demand PS service areas. This is followed by Section IV, where we discuss our findings and their implications, and Section V, which presents a conclusion and recommended future research.

II. METHODOLOGY

A. Data Sources

The data used for this analysis came from public repositories. The data about public CPs and charging sessions for both SP Energy and SSEN came from the ChargePlace Scotland's online data repository [13]. Data about number of EVs in PS service areas for SP Energy came from the SP Energy Distribution Future Energy Scenario (DFES) forecasts posted online [14], and for SSEN, they came from data used for the DFES forecasts available on the SSEN Distribution Data Portal [15].

Data about PS service areas for SP Energy for 2023 were obtained from SP Energy Networks, and for PS service areas for SSEN, 2023 data were obtained from the SSEN data portal [14][15]. Each dataset included number of EVs in each PS service area, and we connected these data to the CP data to determine the number of public CPs in each PS service area.

B. Analytic Approach

First, public CP data from December 2024 for both SP Energy and SSEN were obtained from ChargePlace Scotland, which included 2,848 CPs in total [13]. CP monthly variables that were analysed included number of charging sessions, power drawn (kWh), number of faults (outages), and percentage of uptime (see Table 1 for categories). A bivariate analysis was conducted on public CP usage for both SP Energy and SSEN in December 2024 on the basis of these variables. Based on available data, we established 5 EVs per public CP in each PS service area as an acceptable public CP density.

To identify areas of potentially high demand for public CPs, we compared maps of both SP Energy and SSEN PS service areas in terms of EV prevalence and CP density. An interactive geospatial map was developed to visualise the distribution of EV counts and associated public CPs across PS service areas. This analysis used Python to integrate postcode-level geolocation data to enable aggregation and enumeration of CP density and EV prevalence by PS service areas. CP locations were spatially joined with PS polygons using a Geographic Information System (GIS) method to determine PS service area inclusion. Distinct colour schemes were applied to PS service areas based on thresholds for both CP density and EV prevalence. The final map was rendered using the Folium Python library, which allows an intuitive, zoomable web-based exploration [16].

TABLE I
CATEGORIES.

Category	Level
Number of Sessions	0 to 5
	6 to 19
	20 to 49
	50 or more
Power Drawn (kilowatt hours)	0 to < 75
	75 to <400
	400 to <1,000
	1,000 or more
Number of Faults	0
	1
	2 or more
Uptime	100%
	<100%
Number of Connectors	1
	2
	3 or more
	Unknown
Session Start Time of Day	Work Day (8:00 to 18:00)
	Evening (18:00 to 22:00)
	Sleep Time (22:00 to 6:00)
	Early AM (6:00 to 8:00)
Duration of Session	<30 minutes
	30 minutes up to 1 hour
	1 hour up to 2 hours
	2 or more hours

It was determined that PS service areas in the SP Energy network likely suffered from high demand, as many PS service areas contained 100 or more EVs, but less than 10 public CPs. As described in Global EV Outlook 2024, a ratio of 100:1 to 10:1 with respect to EVs in a region to public CPs in the region would indicate very low access to public CPs, as this entire ratio range is much higher than most country averages [17]. We classified these areas as "high demand" PS service areas. We also used session data to characterise public CPs as having 1, 2, or 3 or more connectors. We classified public CPs as to whether or not they fell in high demand PS service areas, and conducted a bivariate analysis. To explore this potential for unmet demand for public CPs, we analysed sessions from public CPs in SP Energy PS service areas from July 2024 through December 2024, using session data from ChargePlace Scotland [13] (see Table 1 for categories).

We hypothesised that sessions from public CPs in high demand PS service areas served by SP Energy would be more likely to start outside of work day times. We also hypothesized that sessions from public CPs in high demand PS service areas would show a different duration pattern than those from public CPs in other PS service areas. After conducting a bivariate analysis, to derive a profile including month of session, time of day of session start, and duration of session, for sessions occurring at public CPs in high demand PS service areas, we developed a logistic regression model, including these factors as independent variables predicting whether or not the session occurred at a public CP in a high demand PS service area.

TABLE II
BIVARIATE ANALYSIS OF PUBLIC ELECTRIC VEHICLE CHARGING POINTS IN SCOTTISH POWER ENERGY, HIGH DEMAND VS. OTHER PRIMARY SUBSTATION SERVICE AREAS, DECEMBER 2024.

Category	Level	All n, %	High Demand* n, %	All Others n, %
All	All	1,877, 100%	528, 28%	1,349, 72%
Number of Sessions	0 to 5	469, 25%	106, 20%	363, 27%
	6 to 19	410, 22%	107, 20%	303, 22%
	20 to 49	521, 28%	151, 29%	370, 27%
	50 or more	477, 25%	164, 31%	313, 23%
Power Drawn (kilowatt hours)	0 to < 75	435, 23%	102, 19%	333, 25%
	75 to <400	440, 23%	96, 18%	344, 26%
	400 to <1,000	476, 25%	148, 28%	328, 24%
	1,000 or more	526, 28%	182, 34%	344, 26%
Number of Faults	0	1,317, 70%	354, 67%	963, 71%
	1	318, 17%	103, 20%	215, 16%
	2 or more	242, 13%	71, 13%	171, 13%
Uptime	100%	1,407, 75%	393, 74%	1014, 75%
	<100%	470, 25%	135, 26%	335, 25%
Number of Connectors	1	65, 3%	11, 2%	54, 4%
	2	1,377, 73%	395, 75%	982, 73%
	3 or more	284, 15%	77, 15%	207, 15%
	Unknown	151, 8%	45, 9%	106, 8%

Note: First row has row percentages, and the rest are column percentages. *High demand refers to charging points in primary substation service areas where there are 100 or more electric vehicles but < 10 charging points.

Independent variables were coded as indicator variables, and entered into one model. All statistical analysis was done in R GUI [18] using packages *broom* [19], *dplyr* [20], *ggplot2* [21], *lubridate* [22], *readr* [23], and *scales* [24].

III. RESULTS

Data were available from 2,848 public CPs in SP Energy and SSEN PS service areas. The data that were complete and usable represented 2,810 public CPs which were included in the dataset. Data on 820 PS service areas from both networks were also included in the analysis. For monthly sessions from SP Energy, data from 833,691 sessions available for analysis. From these, 323 were removed due to missing or nonsensical duration values, and 254,074 were removed because they could not be matched to a public CP in a known PS service area, leaving 579,294 sessions that were included in the analysis.

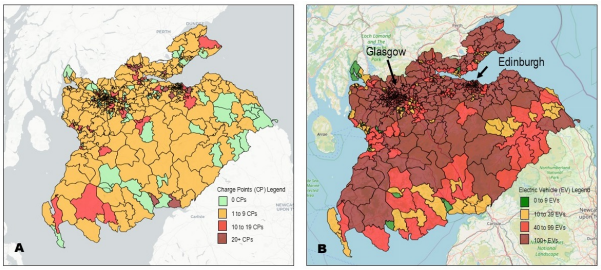

Figure 1. Scottish Power Energy Public Charging Points and Electric Vehicles per Primary Substation Service Areas. Figure 1A indicates public charging point (CP) density, and Figure 1B indicates electric vehicle (EV) density. Major cities are indicated on Figure 1B.

A. Primary Substation Service Areas: EV Ownership and Public CPs

Descriptive statistics were developed with respect to the 820 PS service areas included in the SP Energy and SSEN networks (data not shown). Of the 820 PS service areas analysed, almost half (48%, n = 393) were in the SP Energy network, and the other half (52%, n = 427) were in SSEN. Approximately one third of these PS service areas (34%, n = 279) included 0 public CPs. With respect to number of EVs owned in each PS service area, approximately one fourth (24%, n = 196) contained 0 to 9 EVs, while 27% (n = 223) contained 100 or more EVs. With respect to network, high EV ownership per PS service area was more prevalent in the SP Energy network, compared to SSEN, where 41% (n = 162) of SP Energy PS service areas contained 100 or more EVs, with only 14% (n = 61) of SSEN PS service areas including 100 or more EVs. The distribution of EVs per PS service area was also highly left skewed (median = 36 EVs per PS service area). However, of the 12 PS service areas with 20 or more public CPs, most (n = 10) were in the SP Energy network, with only 2 in the SSEN network.

Approximately two thirds of the 2,810 public CPs analysed in each network fell in the SP Energy network (67%, n = 1,877), with the remainder in SSEN (33%, n = 933). The distribution of number of monthly sessions per public CP was left skewed (median = 21 sessions), but distributions were similar across both networks. A similar distribution pattern was seen in terms of power drawn from public CPs in across networks (median = 484.81 kWh). Overall, most public CPs (70%, n = 1,958) had 0 faults (outages) and maintained 100% uptime (73%, n = 2,045), with a similar distribution for both networks analysed.

B. Comparison of Densities of Public CPs and EV Ownership

Figure 1 shows a comparison of public CP density with EV ownership patterns in SP Energy, and Figure 2 shows the same comparison for SSEN. As depicted in Figure 1, most PS service areas within the SP Energy network contained 100 or more EVs, yet most PS service areas included less than 10 public CPs, implying that there is much higher demand for public CPs than available. In Figure 2, a high prevalence

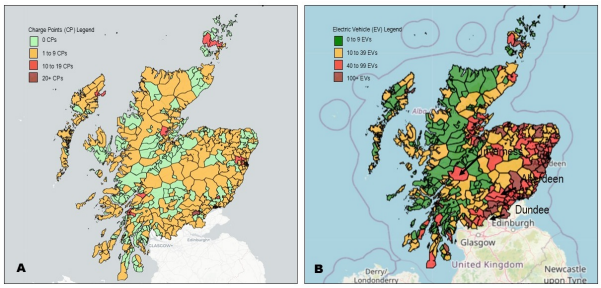

Figure 2. Scottish and Southern Energy Network Public Charging Points and Electric Vehicles per Primary Substation Service Areas. Figure 2A indicates public charging point (CP) density, and Figure 2B indicates electric vehicle (EV) density. Major cities are indicated on Figure 2B.

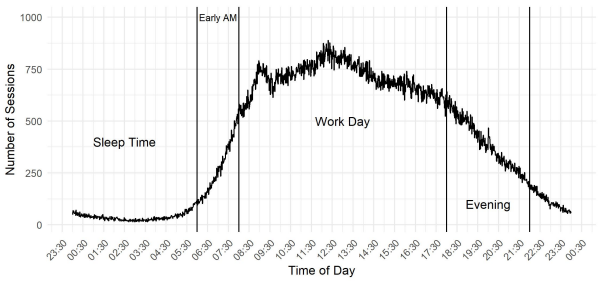

Figure 3. Time Series Plot of Start Time of Day for Sessions from Scottish and Southern Energy Network Public Charging Points, July through December 2024

of EVs in SSEN PS service areas is only evident around the eastern coastal areas, which contain large cities. However, public CP density is not high in these areas, which also may lead to unmet public CP demand. Further, throughout the SSEN network, there were many PS service areas that contain a low prevalence of EVs but 0 public CPs, forcing EV owners to travel to another PS service area to take advantage of public charging services. This implies that while the density of public CPs may be adequate to serve PS service areas in much of the SSEN network, there is also a high probability of unmet demand for public CPs throughout the rest of the network.

C. Charging Session Patterns in High Demand Primary Substation Service Areas

Table 2 presents a bivariate analysis of public CPs in the SP Energy network categorised by those occurring in high demand PS service areas compared to all others. As shown in Table 2, in December 2024, public CPs in high demand PS service areas were more likely to have 50 more sessions than those in other PS service areas (31%, n = 164, compared to 23%, n = 313), with a similar pattern seen in amount of power drawn (1,000 kWh or more, 34%, n = 182 vs. 26%, n = 344). On the other hand, distribution of number of faults, uptime, and number of connectors was similar between high demand PS service areas compared to other service areas.

Figure 3 shows the distribution of all 579,294 charging sessions at public CPs in the SP Energy network over the period of July through December, 2024 by time of day the session started. As displayed in Figure 3, charging sessions were most likely to start during the work day (8:00 to 18:00), reaching peak frequency near 800 sessions, while dropping off during evening hours (18:00 to 22:00) to a frequency of less than 250 sessions. Frequencies of charging session start time were lowest during sleep time (22:00 to 6:00), at times approaching a frequency of 0 charging sessions. During early morning times (6:00 to 8:00), frequencies of charging sessions rose from less than 250 at the beginning of the interval to over 500 at the end of the interval.

Table 3 presents a bivariate analysis of these charging sessions classified by those that took place in high demand PS service areas compared to all others. As shown in Table 3, while most of the sessions (81%, n = 469,681) took place at public CPs outside of high demand PS service areas, monthly distribution of sessions was similar, with the greatest share taking place in August (21%, n = 120,805) compared to the other months analysed. However, a greater percentage of sessions started in evening hours at public CPs in high demand PS service areas compared to others (18%, n = 20,064 compared to 16%, n = 73,339), with a lower percentage starting during the work day (71%, n = 77,975 vs. 74%, n = 347,567). Also, while 37% (n = 40,986) of sessions at public CPs in high demand PS service areas had a duration of 2 or more hours, only 25% (n = 119,533) at public CPs in other PS service areas were 2 or more hours.

Next, a multivariate logistic regression analysis was conducted to develop of a profile of factors associated with sessions taking place at public CPs in the SP Energy network located in high demand PS service areas compared to other PS service areas. The resulting equation was the following:

$$\begin{aligned}\log\left(\frac{p(\text{HD})}{1-p(\text{HD})}\right) &= 0.19110 + 0.8399_{\text{30MIN-1HR}} \\ &+ 1.0675_{\text{1-2HRS}} + 1.6598_{\text{2+HRS}} + 1.1667_{\text{EVE}} \\ &+ 1.0893_{\text{SLEEP}} + 0.9262_{\text{EARLYAM}} + 1.0235_{\text{AUG}} \\ &+ 1.0712_{\text{SEP}} + 1.1040_{\text{OCT}} + 1.0512_{\text{NOV}} \\ &+ 1.0532_{\text{DEC}}\end{aligned} \quad (1)$$

where p(HD) is the probability of the session taking place at a public CP falling in a high demand PS service area (compared to other PS service areas). As indicator variables were used, each level entered into the model had a reference level (<30 minutes for duration, work day for time of day of session start, and July for month of session). In terms of the equation, compared to charging sessions with a duration of less than 30 minutes, sessions had over twice the odds of lasting 30 minutes to 1 hour, almost 3 times the odds of lasting between 1 and 2 hours, and over 5 times the odds of lasting 2 or more hours if they took place at public CPs in high demand PS service areas compared to all other PS service areas. Also, compared to work day times, charging sessions had over 3 times the odds of starting in evening hours, almost 3 times the odds of starting during sleep times, and about 2 and a half

TABLE III
BIVARIATE ANALYSIS OF CHARGING SESSIONS AT PUBLIC ELECTRIC VEHICLE CHARGING POINTS IN SCOTTISH POWER ENERGY, HIGH DEMAND VS. OTHER PRIMARY SUBSTATION SERVICE AREAS, JULY THROUGH DECEMBER 2024.

Category	Level	All	High Demand*	All Others
		n, %	n, %	n, %
All	All Sessions	579,294, 100%	109,613, 19%	469,681, 81%
Month of Session in 2024	July	82,359, 14%	14,930, 14%	67,429, 14%
	August	120,805, 21%	22,519, 21%	98,286, 21%
	September	108,818, 19%	21,146, 19%	87,672, 19%
	October	87,530, 15%	16,647, 15%	70,883, 15%
	November	91,095, 16%	17,407, 16%	73,688, 16%
	December	88,687, 15%	16,964, 15%	71,723, 15%
Session Start Time of Day**	Work Day	425,542, 73%	77,975, 71%	347,567, 74%
	Evening	93,403, 16%	20,064, 18%	73,339, 16%
	Sleep Time	27,599, 5%	5,642, 5%	21,957, 5%
	Early AM	32,750, 6%	5,932, 5%	26,818, 6%
Duration of Session	<30 minutes	170,164, 29%	29,183, 27%	140,981, 30%
	30 minutes up to 1 hour	157,674, 27%	23,141, 21%	134,533, 29%
	1 hour up to 2 hours	90,937, 16%	16,303, 15%	74,634, 16%
	2 or more hours	160,519, 28%	40,986, 37%	119,533, 25%

Note: First row has row percentages, and the rest are column percentages. *High demand refers to charging points in primary substation service areas where there are 100 or more electric vehicles but < 10 charging points. **See Table 1 for exact times.

times the odds of starting in early morning, if they took place at public CPs in high demand PS service areas compared to all other PS service areas. Finally, compared to July, charging sessions from each of the other months analysed had around 3 times the odds of taking place at a public CP in a high demand PS service area compared to all other PS service areas.

IV. DISCUSSION

The findings from the current EV analysis suggest that for the regions served by SP Energy and SSEN, EV uptake has been high, indicating significant progress toward net zero [2]. However, the number of public CPs available to serve EV owners appears to be inadequate, especially in PS service areas served by the SP Energy network. Among the charging sessions taking place in PS service areas in SP Energy, those in areas designated as high demand showed evidence of unmet demand for public CP charging services. Sessions were more likely to start at less convenient times (evening hours, sleep time, and early morning, compared to the work day), and were much more likely to be of extremely long duration (2 or more hours). This was likely because EV owners were competing for use of the few public CPs available in the area, and therefore, were more willing to charge their vehicle at inconvenient times, and were more likely to have their EV stay connected to the charger for a long enough duration to fully charge their vehicle.

This lack of public CPs in the regions served by SP Energy greatly limits the utility of EVs for long trips, and greatly increases range anxiety [9][25–27]. This situation can defeat the purpose of promoting the uptake of EVs, because it can cause EV owners to feel as if they need access to a "back-up" fossil fuel vehicle [25][26]. Expanding the selection of en-route public CPs is a key step in improving charging efficiency for EV owners, and is necessary to alleviate both grid and traffic congestion [28]. A robust strategy for determining the optimal placement and capacity of public CPs is possible to develop through integrating location modelling, with demand forecasting and market penetration analysis [29][30].

The main strength of this analysis is that it uses real-world data to assess the potential for unmet demand for public CPs served by 2 energy networks that contain rural areas thought to be underserved. However, there are also many limitations. Much data had to be removed from the analysis because of missing or nonsensical values, and many assumptions had to be applied to facilitate interpretation.

V. CONCLUSION AND FUTURE WORK

Our research aimed to characterise the distribution of public CPs and EVs in Scotland in the regions served by SP Energy and SSEN, and estimate whether there was unmet demand for public CPs in certain PS service areas. We compared the level of EV ownership with the density of public CPs, and characterised high demand PS service areas as having a high level of EV ownership coupled with a low density of public CPs. We found that the regions served by SP Energy and SSEN had several blackspots which were more prevalent in the SP Energy network, with evidence of unmet demand for access to the public EV charging infrastructure. Sessions in high demand areas were more likely to start at inconvenient times and last a longer duration. Policymakers should focus on ensuring CPs in areas with a high density of EVs have good coverage, and offer more and faster charging connectors. While this analysis has strengths, the results apply specifically to the region analysed. Additionally, the analysis was limited to the use of the variables in public data available, where reliable information about the efficiency of connectors was not available.

In terms of future research, a helpful step for the UK to take to address these blackspots would be to conduct an updated survey of EV owners to better understand their need for public CPs, and set an optimal target for public CP density. These efforts could involve input from Charging Point Network Operators (CPNOs) and Distribution Network

Operators (DNOs), with results reported by PS service areas. Through this collaboration, informed goals could be set for optimal public CP density for each PS service area. In this way, working with existing CPNOs and DNOs, the UK government could ensure that increasing EV ownership across the various energy networks continues to be supported, and effectively facilitate further progression toward their stated net zero goals.

REFERENCES

[1] G. Colverd, R. Bardhan, and J. Cullen, "Benchmarking domestic energy consumption using high-resolution neighbourhood energy data and city clustering in the uk," in *Proceedings of the 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, New York, NY, USA: Association for Computing Machinery, October 2024, pp. 121–131. DOI: 10.1145/3671127.3698183.

[2] HM Government, "Net zero strategy: Build back greener," Controller of Her Majesty's Stationery Office, Tech. Rep., 2022. [Online]. Available: https://assets.publishing.service.gov.uk/media/6194dfa4d3bf7f0555071b1b/net-zero-strategy-beis.pdf, retrieved: January, 2026.

[3] S. Platt and D. Ward, "Hope Valley energy and renewables," Hope Valley Climate Action Energy Group, Tech. Rep. ISBN 978-1-5286-2938-6, 2021. [Online]. Available: https://hopevalleyclimateaction.org.uk/wp-content/uploads/2023/06/Hope-Valley-Energy-and-Renewables.pdf, retrieved: January, 2026.

[4] Y. Wang, J. Wang, and W. He, "Development of efficient, flexible and affordable heat pumps for supporting heat and power decarbonisation in the UK and beyond: Review and perspectives," *Renew. Sustain. Energy Rev.*, vol. 154, pp. 1–25, 2022.

[5] J. Nieto, P. E. Brockway, M. Sakai, and J. Barrett, "Assessing the energy and socio-macroeconomic impacts of the EV transition: A UK case study 2020–2050," *Appl. Energy*, vol. 370, pp. 1–15, 2024.

[6] I. S. Bayram, A. Saad, R. Sims, A. Babu, C. Edmunds, and S. Galloway, "Statistical characterization of public AC EV chargers in the UK," *IEEE Access*, vol. 11, pp. 70 274–70 287, 2024.

[7] C. Skidmore, "Mission zero: Independent review of Net Zero," Net Zero Review Team, Tech. Rep., 2023. [Online]. Available: https://www.gov.uk/government/publications/review-of-net-zero, retrieved: January, 2026.

[8] M. Kumar, K. P. Panda, R. T. Naayagi, R. Thakur, and G. Panda, "Comprehensive review of electric vehicle technology and its impacts: Detailed investigation of charging infrastructure, power management, and control techniques," *Appl. Sci.*, vol. 13, no. 15, pp. 1–58, 2023.

[9] K. Chamberlain and S. A. Majeed, "Evaluating the barrier effects of charge point trauma on UK electric vehicle growth," *World Electr. Veh. J.*, vol. 12, no. 3, pp. 1–40, 2021.

[10] Scottish and Southern Energy Networks, *About Us*, https://www.ssen.co.uk/, retrieved: January, 2026.

[11] SP Energy Networks, *About Us*, https://www.spenergynetworks.co.uk/pages/about_us.aspx, retrieved: January, 2026.

[12] Y. Zhou, C. Essayeh, and T. Morstyn, "Datasets of Great Britain primary substations integrated with household heating information," *Data Brief*, vol. 54, pp. 1–16, 2024.

[13] Charge Place Scotland, *Monthly charge point performance*, https://chargeplacescotland.org/monthly-charge-point-performance/, retrieved: January, 2026.

[14] SP Energy Networks, *SPD DFES totals by substation*, https://spenergynetworks.opendatasoft.com/explore/dataset/spd-dfes-site-forecasts/, retrieved: January, 2026.

[15] SSEN Distribution Data Portal, *Distribution Future Energy Scenarios datasets*, https://data.ssen.co.uk/ssen-distribution/low_carbon_technologies, retrieved: January, 2026.

[16] Folium, *python-visualization*, https://python-visualization.github.io/folium/, 2020, retrieved: January, 2026.

[17] International Energy Agency, "Global EV outlook 2024: Moving toward increased affordability," International Energy Agency, Tech. Rep., 2024. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2024, retrieved: September, 2025.

[18] R Core Team, *R: A Language and Environment for Statistical Computing*, Version 4.5.0 (2025-04-11 ucrt), Data analysis software. Available: https://www.R-project.org/, 2025, retrieved: November, 2025.

[19] D. Robinson et al., *broom: Convert statistical objects into tidy tibbles*, Version 1.0.9, R package. Available: https://cran.r-project.org/web/packages/broom/index.html, March 2025, retrieved: November, 2025.

[20] H. Wickham et al., *dplyr: A grammar of data manipulation*, Version 1.1.4, R package. Available: https://cran.r-project.org/web/packages/dplyr/index.html, November 2023, retrieved: November, 2025.

[21] H. Wickham et al., *Create elegant data visualisations using the grammar of graphics*, Version 3.5.2, R package. Available: https://cran.r-project.org/web/packages/ggplot2/index.html, May 2025, retrieved: November, 2025.

[22] V. Spinu et al., *lubridate: Make dealing with dates a little easier*, Version 1.9.4, R package. Available: https://cran.r-project.org/web/packages/lubridate/index.html, December 2024, retrieved: November, 2025.

[23] H. Wickham et al., *readr: Read rectangular text data*, Version 2.1.5, https://readr.tidyverse.org/, January 2024, retrieved: November, 2025.

[24] H. Wickham et al., *scales: Scale functions for visualization*, Version 1.4.0, https://scales.r-lib.org/, April 2025, retrieved: November, 2025.

[25] K. C. Akshay, G. H. Grace, K. Gunasekaran, and R. Samikannu, "Power consumption prediction for electric

vehicle charging stations and forecasting income," *Sci. Rep.*, vol. 14, no. 1, pp. 1–25, 2024.

[26] H. Budnitz, T. Meelen, and T. Schwanen, "Public residential charging of electric vehicles: An exploration of UK user preferences," *Eur. Transp. Stud.*, vol. 1, pp. 1–12, 2024.

[27] T. Chen et al., "A review on electric vehicle charging infrastructure development in the UK," *J. Mod. Power Syst. Clean Energy*, vol. 8, no. 2, pp. 193–205, 2020.

[28] C. Sommer and M. J. Hossain, "Artificial intelligence-driven optimization of V2G and charging point selection en-route: A systematic literature review," *Energy Convers. and Manag.: X*, vol. 26, pp. 1–22, 2025.

[29] S. Çelik and Ş. Ok, "Electric vehicle charging stations: Model, algorithm, simulation, location, and capacity planning," *Heliyon*, vol. 10, no. 7, pp. 1–19, 2024.

[30] Y. Zhang et al., "Analysis of electric vehicle charging behaviour in existing regional public and workplace charging infrastructure: A case study in the North-East UK," *Transport. Eng.*, vol. 19, pp. 1–13, 2025.

Federated Analytics Hybrid Architecture for Connected Truck Data Analysis

Aslihan Reyhanoglu[1], Feyzi Ege Kumec[1], Abdulkadir Bilge[1],
Hira Bakirhan[2], Bugra Turan[3], and Sinem Coleri [3]
[1]Koc University Ford Otosan Automotive Technologies Laboratory (KUFOTAL), Sariyer, Istanbul, Turkey, 34450
[2]Ford Otosan, Sancaktepe, Istanbul, Turkey, 34885
[3]Department of Electrical and Electronics Engineering, Koc University, Sariyer, Istanbul, Turkey, 34450
E-mail: [areyhanoglu, fkumec, abilge20, buturan, scoleri] @ku.edu.tr and hertugru@ford.com.tr

***Abstract*—Connected vehicles offer opportunities to optimize fleet operations by leveraging data from their on-board sensors. Off-board cloud architectures play a key role in enhancing fleet-wide health monitoring and operational efficiency. Federated Analytics (FA) empowers connected vehicles to enable decentralized data analysis by improving privacy, decreasing cloud data transmission, and supporting real-time decision-making. This paper contributes a practically deployable federated analytics-assisted edge–cloud architecture for connected truck health monitoring, where high-rate telemetry is processed locally on an NVIDIA Jetson Orin and only compact event summaries are transmitted to the cloud for fleet-level reasoning. We further introduce a cloud-side Temporal-Spatial Aggregation (TSA) mechanism that computes near real-time anomaly hotspots using 15-minute sliding windows over geo-fenced zones, enabling explainable time- and location-correlated insights. Finally, we demonstrate feasibility with real truck data and provide an End-to-End (E2E) latency and Uplink (UL) data volume reduction against a raw data transfer baseline.**

***Keywords*— *Federated Analytics*; *Edge Computing*; *Connected Trucks*; *Temporal–Spatial Aggregation*; *Fleet Health Monitoring*.**

I. Introduction

Connected vehicles offer opportunities to enhance safety and performance using data-driven insights from on-board sensors for fleet management. Complex data and heavy transmission demands can overwhelm centralized systems, leading to high latency, network congestion, and costly cloud storage. 5G Automotive Association (5GAA) defines for remote vehicle health monitoring, a latency of less than 30 seconds [1] is acceptable, allowing periodic updates without compromising safety. On the other hand, 5GAA [2] specifies a latency of 100 milliseconds to support data sharing use cases of dynamic objects. To address these latency requirements, technologies, such as FA and Multi-Access Edge Computing (MEC) are provisioned to be integrated into Vehicle-to-Network-to-Vehicle (V2N2V) schemes [3]. V2N2V systems enable seamless communication between vehicles and the cloud, supporting real-time decision-making [4]. V2N2V use cases aim to support cooperative collision avoidance and driving, together with fleet management use cases, requiring low latency and high reliability [5]. Unlike Federated Learning (FL), which primarily exchanges model updates to collaboratively train a global model, FA focuses on decentralized data analysis where raw data remains on the vehicles and only compact analytical outputs (events/aggregates) are shared. FA addresses latency, scalability, and privacy challenges in connected vehicle systems by combining localized edge computing with cloud-based aggregated data analysis [6]. Edge computing enables real-time event detection whereas reducing cloud dependency, latency, and resource usage.

In the literature, significant advancements have been proposed to improve the reliability, latency, and efficiency of connected vehicle systems through V2N2V communications [5], federated analytics [6]–[8], and edge computing [9], [10]. In [5], an analytical model incorporating numerology, scheduling, and Hybrid Automatic Repeat Request (HARQ) evaluates 5G New Radio (NR) V2N2V configurations to meet the 6 ms UL+Downlink (DL) radio-latency budget derived from the 3rd Generation Partnership Project (3GPP) 10 ms E2E target for High Level of Automation (HLoA) cooperative lane change. The decentralized federated analytics framework in [6] converges with fewer communication rounds and reduces communication cost by up to 89% in a simulated Vehicular Ad Hoc Network (VANET) with Vehicle-to-Vehicle (V2V) neighbor broadcast. Similarly, [7] introduces a federated anomaly detection approach that filters malicious updates without centralized data, achieving up to 6.9× higher accuracy under strong attacks and offering formal robustness guarantees. In [8], a privacy-preserving framework enables vehicles to transmit locally processed analytics insights, such as object detection results to nearby MEC/Road Side Unit (RSU)s. These edge nodes aggregate the data and broadcast updated High Definition (HD) map tiles. A two-tier 5G edge infrastructure, optimized using queuing theory, achieves optimized latency for map updates in dense urban environments, as demonstrated through Simulation of Urban MObility (SUMO) simulations. [9] shows that in realistic multi-Mobile Network Operator (MNO) 5G V2N2V scenarios, edge-proximal MEC deployments with local peering can keep the average E2E latency below 20 ms for a wide range of loads, indicating feasibility for time-critical cooperative maneuvers such as lane merging. [10] shows that MEC federation can reduce E2E latency by over 80% in multi-MNO 5G V2N2V scenarios, supporting the feasibility of sub-20 ms cooperative maneuvers such as lane merging. However, none of the studies to date addressed real-time truck health monitoring using processed anomalies without heavy on-board retraining. By combining low-latency edge detection with cloud-based temporal–spatial aggregation, we offer a scalable solution for fleet-wide monitoring.

This paper presents a FA-assisted edge–cloud architecture for connected trucks, where raw sensor data is processed on NVIDIA Jetson Orin device to generate lightweight event

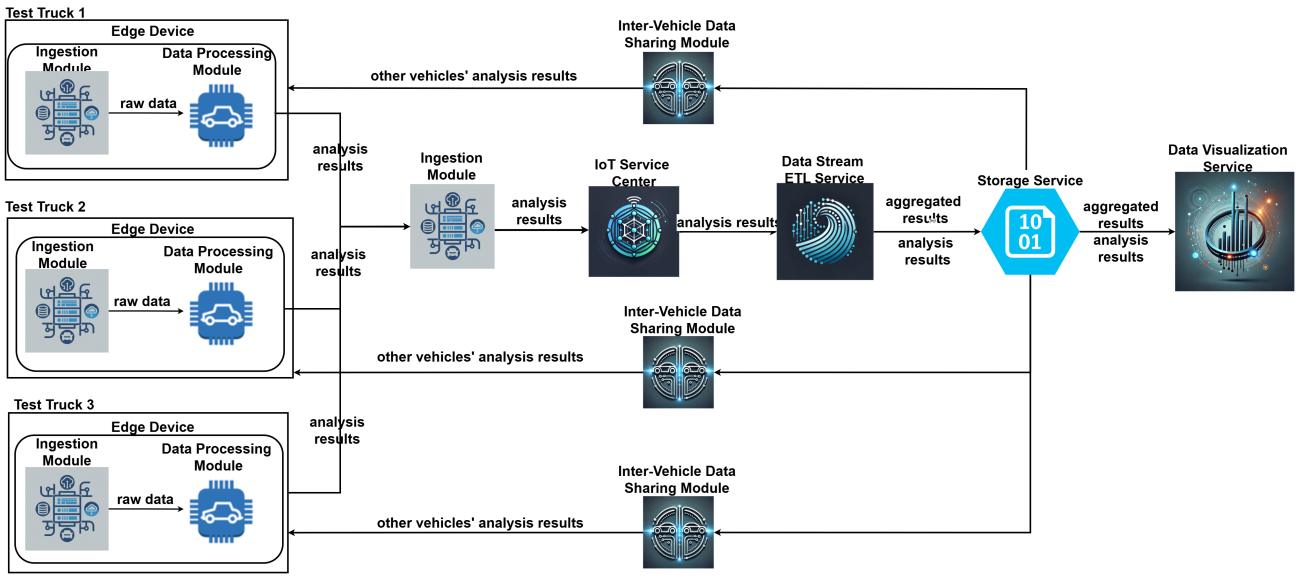

Figure 1. Proposed Federated Analytics Architecture (FA Architecture).

summaries for real-time truck health monitoring. We also introduce cloud-based TSA, using 15-minute time windows over geo-fenced zones to create real-time anomaly hotspot maps. We demonstrate efficient data reduction and achieve approximately 1-second E2E latency by using real data from Ford F-Max test trucks and synthetic data.

The rest of the paper is organized as follows. Section II presents the system model. Section III describes the algorithms for analysis and aggregation. Section IV provides benchmarking and performance evaluation results, focusing on end-to-end latency metrics and comparisons with the raw data transfer baseline. Finally, Section V concludes the paper and outlines future work.

II. System Model

This section presents the proposed FA Architecture (see Figure 1) for scalable, secure, and real-time truck behavior and health monitoring in connected fleet systems. It supports continuous data collection, processing, storage, and visualization across on-board and off-board systems. The architecture includes four main components: the Ingestion Module, Data Processing Module, Off-Board Backend, and Inter-Vehicle Data Sharing Module. On-board modules run on an NVIDIA Jetson Orin System-on-Chip (SoC), enabling real-time edge processing. The Off-Board Backend handles long-term storage, data exchange, and visualization through its Internet of Things (IoT) Service Center, Extract-Transform-Load (ETL) Service, Storage Service, and Visualization Service. Table I summarizes the modules and their relationship to FA.

The Ingestion Module (see Figure 2) acts as the communication bridge between edge devices and the cloud. It collects raw sensor data via a Robot Operating System (ROS) topic consumer and forwards these messages to the on-board Data Processing Module. In addition, it transmits the processed results to the cloud in real time, enabling bidirectional communication and

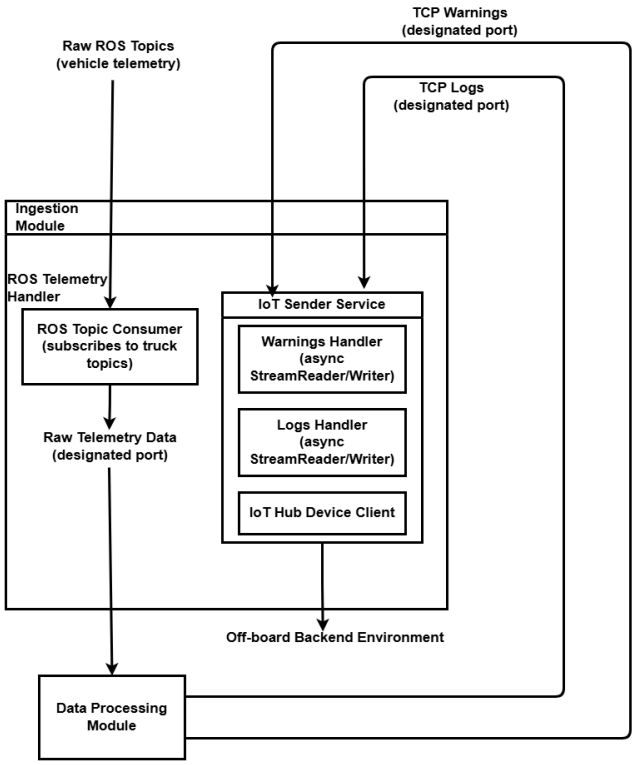

Figure 2. Architecture of the ingestion module.

scalable connection management. Each truck is registered as a distinct IoT entity, which enhances security and access control.

The Data Processing Module serves as the on-board analytics engine, processing high-frequency telemetry to detect anomalies, such as harsh braking, aggressive acceleration, steering issues, and sensor faults. It uses socket connections to stream data and applies thresholds derived from domain-specific

heuristics to identify abnormal behavior.

The IoT Service Center enables bidirectional communication by facilitating devices to transmit telemetry data to the cloud and also receive updates and commands from the cloud [11]. In our context, it plays a role as a secure and scalable bridge between data transfer module and cloud services. We create three separate devices on the IoT Service Center to manage and monitor each vehicle's data individually. Each vehicle has its own identity within the IoT system, which enhances organization, security, and scalability in the IoT architecture.

The Data Stream ETL Service aggregates telemetry and anomaly events (e.g., throttle, brake, GPS, steering, sensor health) using 15-minute sliding windows with 1-minute hops. It streams edge-processed insights in real time to the Storage Service [12].

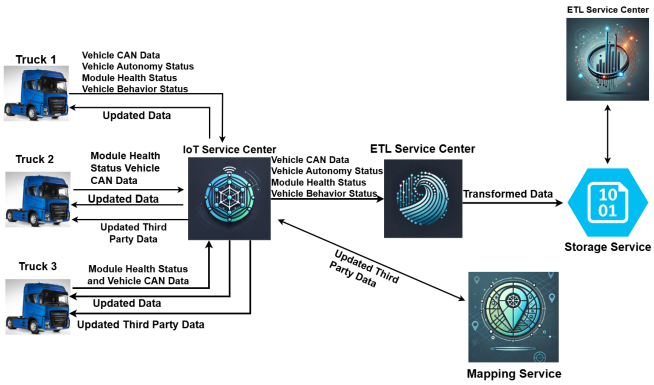

Figure 3. An Architecture for Raw Data Transfer.

The Storage Service provides long-term storage for analysis results and aggregated data, organized hierarchically by vehicle and timestamp. This structure improves retrieval efficiency and reduces operational costs [13].

The Data Visualization Service enables real-time analytics and querying through external tables, avoiding the need for data import. Dashboards show module health, warnings, mode changes, and system metrics for both real-time and historical analysis.

The Inter-Vehicle Data Sharing Module allows connected trucks to exchange processed data in real time. It retrieves and processes only the latest data (past 60 seconds) to ensure relevance and uses parallel processing with log rotation to manage system load.

To protect fleet data, the architecture uses layered security: isolated Docker networks, Transport Layer Security (TLS) encryption via an Nginx reverse proxy, and strict routing rules. Only flagged events with minimal metadata (timestamp, location) are sent to the cloud, ensuring privacy and reducing system load.

Figure 3 shows the baseline architecture, where raw data is sent directly to the cloud without edge processing. This model is used to benchmark the improved performance and latency of the proposed FA Architecture.

III. Algorithms for Analysis and Aggregation

This section outlines the core algorithms implemented in our FA architecture, including both the on-board Data Processor Algorithm (See Figure 4) and the off-board Temporal–Spatial ETL Streaming Algorithm (See Figure 5).

The Data Processor Algorithm runs independently on each vehicle to detect abnormal driving behavior and system anomalies in real time. Threshold values are informed by domain-specific heuristics and common telematics practices, and are applied via configurable rule checks to ensure deterministic real-time operation. A background daemon loads configurations—network endpoints, rule checks, and thresholds—and listens for Transmission Control Protocol (TCP) connections. Each connection is handled by a dedicated thread that parses JSON-formatted telemetry data into a queue. A worker thread pool processes this data using rule-based checks to detect events such as: aggressive acceleration (throttle > 90% for > 15s), harsh braking (brake > 80% more than 3 times in 5 mins), erratic steering (angle > 30° at least 3 times in 10s), and lane departure risks (yaw error > 0.05 rad or lateral deviation > 0.5 m). It also monitors module health—triggering alerts if any sensor (camera, radar, traffic predictor) reports faults in 10 consecutive readings over 20s—and flags abrupt control transitions (mode changes within 2s). Triggered events are de-duplicated within 5 seconds and encoded into a compact JSON with vehicle ID (i), warning type (r), and event hits (h: timestamps and GPS), enabling efficient, low-bandwidth reporting without raw data transfer.

The Temporal–Spatial ETL Streaming Algorithm runs off-board and aggregates processed warnings in real time. Each event is mapped to a geo-fenced zone based on GPS coordinates, representing regions like delivery routes or test areas. A sliding 15-minute time window, updated every minute, groups events by zone and computes two metrics: the total number of threshold-exceeding events and the number of unique vehicles involved. For example, in brake analysis, it counts high-braking events and identifies how many different trucks triggered them in a given zone and time frame. The output is a JSON object with fields: z (zone ID), ws and we (window start and end in ISO 8601), hb (high brake events), and av (active vehicles). This enables near real-time visibility into fleet behavior, helping managers detect and localize safety concerns by time and location.

IV. Benchmarking and Performance Evaluation

We provide a comprehensive benchmarking and performance evaluation of the proposed FA Architecture, focusing on key latency metrics to assess system responsiveness. The evaluation includes average E2E latency under varying data rates and network conditions. Real-world data is used to compare the baseline and FA architectures, whereas synthetic data is used for controlled experiments. No data filtering is applied during testing. To ensure a detailed analysis, we break down the latency measurements for the FA Architecture into several key components. First, the On-board to Off-board latency refers to

TABLE I.
MODULES AND THEIR ROLES WITHIN THE PROPOSED FEDERATED ANALYTICS (FA) ARCHITECTURE.

Module	Location & FA Role	Key Function & Output Format
Ingestion Module	On-board (Jetson Orin) ↔ Cloud (supporting)	Collects ROS telemetry and forwards messages to the on-board *Data Processing Module*. Transmits processed data to the cloud; outputs `Telemetry JSON`.
FA Processor (Data Processing Module)	On-board (Jetson Orin) **Core FA module**	Performs real-time anomaly detection with rule-based thresholds; outputs compact `Warning JSON`.
ETL / TSA Streaming	Cloud (post-FA aggregation)	Aggregates warnings in 15-min sliding windows; outputs `Hotspot Summary JSON`.
Storage Service	Cloud	Hierarchical blob storage for long-term archival and retrieval.
Visualization Service	Cloud (interface)	Real-time dashboards and historical analytics for fleet managers.
Inter-Vehicle Data Sharing	On-board (Jetson Orin) ↔ Cloud (FA-enabled sync)	Distributes last 60 s of FA warnings to peer vehicles; outputs `Recent Warnings JSON`.

the delay experienced during the transmission of data from the vehicle to the off-board backend environment. Next, the IoT Service Center to Data Stream ETL latency captures the time taken from when data is enqueued in the IoT Service Center until it is processed by the Data Stream ETL Service. Finally, the Average E2E latency (Vehicle-to-Cloud-to-Vehicle) represents the total time taken for data to travel from the originating vehicle to the cloud, undergo processing, and be transmitted to another connected vehicle. This final metric encompasses all stages of data transmission, processing, and response delivery. We evaluate the system's performance under two distinct experimental setups.

In the first setup, we compare the performance of the proposed Federated Analytics (FA) Architecture with a baseline architecture using real-world data from instrumented Ford F-Max test trucks. In the baseline, raw data is transmitted directly between vehicles via the IoT Service Center without any intermediate analytics. In the FA Architecture, Vehicle 1 performs local analysis on an on-board NVIDIA Jetson Orin device, and only the results are transmitted to the cloud. Vehicle 2 then retrieves the aggregated output from the Storage Service. All tests use unidirectional communication and are repeated five times, with the mean values used for evaluation. The experiment involves three types of data streams: module status (20Hz), Highway Pilot (HP) vehicle status (100Hz), and a combined stream. These are transmitted through the Ingestion Module, which includes a ROS topic consumer for real-time handling. The module status stream reports health indicators, whereas the HP vehicle status stream includes data, such as steering angle, throttle input, and engine signals. Raw message sizes are approximately 180 bytes for module status and 1850 bytes for HP vehicle status. In contrast, the FA Architecture reduces data size by transmitting analytics outputs—generated by the Data Stream ETL Service—as compact JSON messages of up to 200 bytes. These include warnings like steering anomalies and module faults, significantly lowering transmission and storage costs. This reduction is most pronounced for high-frequency streams like HP vehicle status, whereas gains for lower-bandwidth streams like module status are more modest, as event summaries are already compact by nature.

TABLE II.
LATENCY COMPARISON FOR RAW DATA TRANSFER AND FA ARCHITECTURES.

Topics	Raw (s)	FA (s)
module_status (20 Hz)	0.55	1.03
hpvehiclestatus (100 Hz)	0.62	1.11
Combined Streams	0.746	1.12

Table II compares average end-to-end latency between the raw data transfer architecture and the proposed FA Architecture. For the module_status stream at 20Hz, latency increases from 0.55 seconds (raw) to 1.03 seconds (FA), mainly due to the overhead of retrieving processed data from Blob Storage. A similar pattern is seen with the HP vehicle status stream at 100Hz, where latency rises from 0.62 seconds to 1.11 seconds. For combined streams, the raw architecture yields an average latency of 0.746 seconds, compared to 1.12 seconds under the FA Architecture.

In the second setup, we assess the performance of the proposed FA Architecture under varying data frequencies using synthetic data in a bidirectional communication setup. Both devices process, transmit, and receive information simultaneously. Structured synthetic telemetry messages, each around 3,311 bytes, are generated and sent to test system behavior under different data rates. Each test iteration includes one AUTO and one MANUAL transition message to simulate abrupt control mode changes. To measure end-to-end latency, timestamps are embedded in the JSON analysis results produced by the on-board module, while additional metadata is appended on the backend. The maximum size of analysis results retrieved from blob storage during these tests is 134 bytes.

Table III shows the latency performance of the proposed FA Architecture in a bidirectional setup under varying data frequencies. One instance runs on an NVIDIA Jetson Orin (Vehicle 1), the other on a laptop (Vehicle 2). The measurements

```
Require: Environment variables (hosts, ports, flags, thresholds)
Ensure : Running data-processor service
processor ← DataProcessor(env.vehicle_id) open TCP socket on (listen_host, listen_port)
while true do
    (conn, addr) ← accept() spawn client_handler(conn, processor)
end
Procedure client_handler(conn, processor):
    foreach line in conn stream do
        row ← parseJSON(line) processor.enqueue_data(row)
    end
Class DataProcessor(id):
    start log-worker, processing-worker, warning-sender threads
    Procedure enqueue_data(row):
        processing_queue.put(row)
    Procedure processing_worker():
        while true do
            row ← processing_queue.get() if row is shutdown then
                return
            end
            analyze(row)
        end
    Procedure analyze(row):
        purge_old_data() update_history(row) foreach check in {throttle, brake, gps, steering, risky, modules} do
            if check.enabled then
                run_check(row)
            end
        end
    Procedure enqueue_warning(text):
        if not duplicate in last 5 s then
            warning_queue.put(text)
        end
    Procedure warning_sender():
        while true do
            w ← warning_queue.get() if w is shutdown then
                return
            end
            send_over_tcp(w)
        end
```

Figure 4. Data Processor Algorithm.

TABLE III.
LATENCY ANALYSIS UNDER DIFFERENT DATA FREQUENCIES FOR 10-MINUTE DATA FLOW BETWEEN VEHICLES.

Data Freq.	V1 → V2			V2 → V1		
	On→Off	IoT→ETL	Avg. E2E	On→Off	IoT→ETL	Avg. E2E
0.5 Hz	0.083s	0.107s	1.18s	0.009s	0.109s	1.07s
1 Hz	0.018s	0.108s	1.05s	0.009s	0.104s	0.87s
5 Hz	0.008s	0.110s	0.76s	0.008s	0.104s	0.76s
10 Hz	0.009s	0.105s	0.80s	0.008s	0.104s	0.79s
20 Hz	0.009s	0.109s	0.86s	0.128s	0.114s	1.10s

include on-board to off-board latency, IoT service to data stream ETL latency, and average E2E latency in both directions. As frequency increases from 0.5Hz to 10Hz, latency values decrease and stabilize, with optimal performance observed between 5Hz and 10Hz. In this range, E2E latency remains between 0.76s and 0.80s in both directions. At 20Hz, latency increases due to higher computational load and queuing. On-board to off-board latency remains stable at 0.009s (Vehicle 1 to 2) and rises to 0.128s (Vehicle 2 to 1). Average E2E latency grows to 0.86s and 1.10s, respectively. IoT service to ETL latency remains steady between 0.104s and 0.110s up to 10Hz,

```
Require: Continuous stream of telemetry events
Ensure : Real-time aggregates pushed to Storage Service
// Main ingestion loop
while event ← ingest_next() do
    event.timestamp ← extract_event_time(event) if not is_anomaly(event) then
        continue
    end
    event.zone ← lookup_geo_fence(event.lat, event.lon) aggregate_queue.put(event)
end
Procedure aggregate_worker()
    while true do
        window ← new_window(15 min, 1 min) buffer ← dequeue_events(window)
        foreach zone in buffer.group_by(zone) do
            events_z ← buffer.filter(zone) hb ← count_if(events_z, event.value ¿ threshold) av ← distinct_count(events_z, event.vehicle_id) result ← { "z": zone, "ws": window.start, "we": window.end, "hb": hb, "av": av }
            send_to_storage(result)
        end
        sleep_until(next_hop_time())
    end
end
```

Figure 5. Temporal-Spatial ETL Streaming Analysis Algorithm.

with only a slight increase at 20Hz. Furthermore, E2E latency from Vehicle 1 to 2 rises by 13.16% (0.76s to 0.86s), while latency from Vehicle 2 to 1 increases by 44.74% (0.76s to 1.10s). These results highlight the trade-off between data frequency, processing capability, and network performance, stressing the importance of tuning the system for high-frequency workloads.

Beyond latency, the system supports operational effectiveness by producing explainable event warnings and enabling TSA-based hotspot identification for time- and location-correlated insights. Robustness is supported by short-term de-duplication and consecutive fault detection, which help filter transient noise. Finally, scalability is enabled by the event-driven architecture and fixed-window aggregation, which largely decouples cloud processing and uplink costs from raw telemetry volume.

V. Conclusions and Future Work

This paper presents a federated analytics-assisted platform that combines edge computing and cloud services to support real-time monitoring in connected truck fleets. Telemetry data is processed locally on Jetson Orin devices using domain-specific threshold checks to reduce data transmission, preserve privacy, and ensure low-latency responses. Aggregated insights are generated through spatio-temporal analysis in the cloud, and visualized via interactive dashboards for fleet-wide operational awareness. Experimental results show that the system maintains an average E2E latency of approximately 1 second even at 20 Hz data rates, demonstrating scalability and efficiency. Future work will focus on latency optimization through edge-cloud synchronization and 5G integration. By leveraging local processing capabilities, we will also aim to implement lightweight unsupervised learning models that can dynamically adjust thresholds based on individual driver profiles and environmental contexts, further enhancing the robustness of fleet management solutions.

References

[1] 5G Automotive Association, "C-V2X Use Cases, Methodology, Examples and Service Level Requirements," 5G Automotive Association, White Paper, June 2019.

[2] 5G Automotive Association, "Use Case Implementations for Sensor Data Sharing," 5G Automotive Association, White Paper, 2023.

[3] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1628–1656, 2017.

[4] M. C. Lucas-Estañ *et al.*, "V2x service provisioning with 5g v2n2v communications with cross-stakeholder information sharing," in *2024 IEEE Vehicular Networking Conference (VNC)*, 2024, pp. 109–112.

[5] M. C. Lucas-Estañ *et al.*, "An analytical latency model and evaluation of the capacity of 5g nr to support v2x services using v2n2v communications," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 2, pp. 2293–2306, 2023.

[6] L. Zhao *et al.*, "A decentralized communication-efficient federated analytics framework for connected vehicles," *IEEE Transactions on Vehicular Technology*, 2024.

[7] S. Shi, C. Hu, D. Wang, Y. Zhu, and Z. Han, "Federated anomaly analytics for local model poisoning attack," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 2, pp. 596–610, 2022.

[8] L. Toka *et al.*, "5g on the roads: Latency-optimized federated analytics in the vehicular edge," *IEEE Access*, vol. 11, pp. 81 737–81 752, 2023.

[9] B. Coll-Perales *et al.*, "End-to-end latency of v2n2v communications under different 5g and computing deployments in multi-mno scenarios," in *2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2022, pp. 622–627.

[10] B. Coll-Perales *et al.*, "Improving the latency of 5g v2n2v communications in multi-mno scenarios using mec federation," in *2022 IEEE 95th Vehicular Technology Conference (VTC2022-Spring)*, 2022, pp. 1–5.

[11] Y. Liu, K. A. Hassan, M. Karlsson, Z. Pang, and S. Gong, "A data-centric internet of things framework based on azure cloud," *IEEE Access*, vol. 7, pp. 53 839–53 858, 2019.

[12] R. Dautov and S. Distefano, "Stream processing on clustered edge devices," *IEEE Transactions on Cloud Computing*, vol. 10, no. 2, pp. 885–898, 2022.

[13] M. Waly, *Learning Microsoft Azure Storage: Build Large-Scale, Real-World Apps by Effectively Planning, Deploying, and Implementing Azure Storage Solutions*. Packt Publishing Ltd, 2017.

Clustering-Based Anomaly Detection for Connected Trucks

Abdulkadir Bilge[1], Erdem Ergen[1], Onur Ciliz[1], Aslihan Reyhanoglu[1], Feyzi Ege Kumec[1],
Hira Bakirhan[2], Bugra Turan[3], and Sinem Coleri [3]
[1]Koc University Ford Otosan Automotive Technologies Laboratory (KUFOTAL), Sariyer, Istanbul, Turkey, 34450
[2]Ford Otosan, Sancaktepe, Istanbul, Turkey, 34885
[3]Department of Electrical and Electronics Engineering, Koc University, Sariyer, Istanbul, Turkey, 34450
E-mail: [abilge20, eergen20, ociliz21, areyhanoglu, fkumec, buturan, scoleri] @ku.edu.tr and hertugru@ford.com.tr

***Abstract*—Proactive fault detection and anomaly detection are among key enablers of autonomous driving technologies provisioned to reduce operational downtime and increase reliability. Unsupervised learning methods are utilized to detect patterns in large datasets for clustering. In this study, we demonstrate clustering-based anomaly detection using engine-related onboard telemetry from connected trucks. First, we prepare the data by missing value removal, relevant feature selection, and data standardization in the preprocessing step. Then, we apply five clustering methods: K-Means, Isolation Forest, Z-Score Analysis, Gaussian Mixture Model (GMM) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Finally, the silhouette score, Calinski-Harabasz index, and domain-specific thresholds are employed to validate the models' performance in anomaly detection. The results demonstrate that the best performing clustering method is Z-Score.**

***Keywords*— *Connected trucks*; *anomaly detection*; *unsupervised learning*; *clustering*; *vehicle telemetry*.**

I. Introduction

Anomaly detection in connected trucks is essential for predictive maintenance and enhanced fleet efficiency, leveraging vehicle diagnostics, geolocation, driving behavior, and weather data to extract actionable insights despite large, complex datasets. Machine learning models predict maintenance needs, detect anomalies, and optimize operations for safer, more cost-effective fleet management [1]–[3].

In the automotive industry, corrective maintenance is employed to fix component failure [4]. On the other hand, Predictive Maintenance (PdM) is a novel concept that focuses on predicting potential vehicle failures before they occur by analyzing vehicle data [5]. PdM reduces unplanned downtime, leading to fewer unexpected repairs, increasing efficiency and user availability. Unsupervised learning techniques offer robust and effective anomaly detection algorithms to handle complex and large-scale datasets in the context of connected truck data analysis. They provide an optimal framework for the design of anomaly detection algorithms to identify anomalies in vehicle data, improve predictive maintenance, and optimize fleet efficiency [6]. Unsupervised methods are commonly preferred for pattern recognition and outlier identification [7].

Building on these methodological insights, recent studies have applied unsupervised learning and anomaly detection techniques across various domains within the automotive industry, including driving behavior anomaly detection [8], Electric Vehicle (EV) battery anomaly detection [9], sensor data abnormality monitoring [10], intravehicular communication abnormality [11] and cybersecurity-intrusion protection [12] [13]. The authors in [8] compare the performance of supervised (Support Vector Machine (SVM), k-Nearest Neighbor (KNN)) and unsupervised (isolation forest [14], Local Outlier Factor (LOF), Z-score [15]) approaches for detecting behavioral abnormalities. Their study utilizes Principal Component Analysis (PCA) and the Minimum Covariance Determinant (MCD) methods for dataset analysis. They conclude that the MCD algorithm demonstrated remarkable results regarding the metrics of accuracy, F1-score, and Mean Absolute Error (MAE). [9] introduces an anomaly detection framework for EV batteries to avoid problems such as thermal runaway and overheat, based on time series analysis. They report that their framework outperforms traditional anomaly detection methods, including PCA, KNN and autoencoders. [10] develops a univariate time series data driven anomaly detection model to detect anomalies in time series vehicle sensor data. They conclude that isolation forest executes the best performance for unsupervised anomaly detection on time series vehicle sensor data. [16] proposes a two-stage approach for radio resource management in Cellular Vehicle-to-Everything (C-V2X) networks. In the first stage, DBSCAN is utilized to cluster vehicles based on their geographical locations, predicted positions, and speeds. The study demonstrates that using DBSCAN for vehicle clustering enhances spectral efficiency and optimizes resource allocation in vehicular networks. In [17], the authors propose an innovative anomaly detection algorithm, namely the Long Short Term Memory (LSTM) Autoencoder with GMM, to detect anomalous behavior in Connected and Autonomous Vehicle (CAV) trajectories. Previously mentioned studies lack a comprehensive, multi-algorithm approach that integrates diverse vehicle subsystems, domain-specific thresholds, and robust preprocessing for heterogeneous vehicle sensor data, limiting their effectiveness in general-purpose, real-world predictive maintenance.

This paper benchmarks multiple unsupervised anomaly detection methods on heterogeneous, real-world engine-related connected-truck telemetry within a unified evaluation pipeline. The approach leverages unsupervised learning techniques, specifically K-Means clustering [18], Isolation Forest, Z-score analysis, GMM and DBSCAN to identify abnormal patterns and potential system faults.

The paper is organized as follows: Section II describes the

 ISBN: 978-1-68558-348-4

dataset and preprocessing. Section III presents the anomaly detection algorithms. Section IV provides the performance evaluation and results. Section V concludes the paper with remarks and possible future work.

II. System Model

This section outlines the system model for detecting anomalies in engine-related connected-truck telemetry data.

The dataset used in this study is collected over a period of 14 consecutive days from two identical connected trucks. The entire dataset includes 90 features and 952815 data points. The data is off-boarded from the truck's onboard telemetry system, which captures a comprehensive set of features that reflect its operational behavior and performance. Each feature is recorded with its own specific frequency, creating a heterogeneous dataset requiring careful alignment and synchronization during preprocessing. Data challenges include missing values due to sensor failures and irregular sampling rates stem from features being logged at different intervals. These issues are meticulously addressed during the preprocessing stage to ensure the reliability and suitability of the data set for anomaly detection.

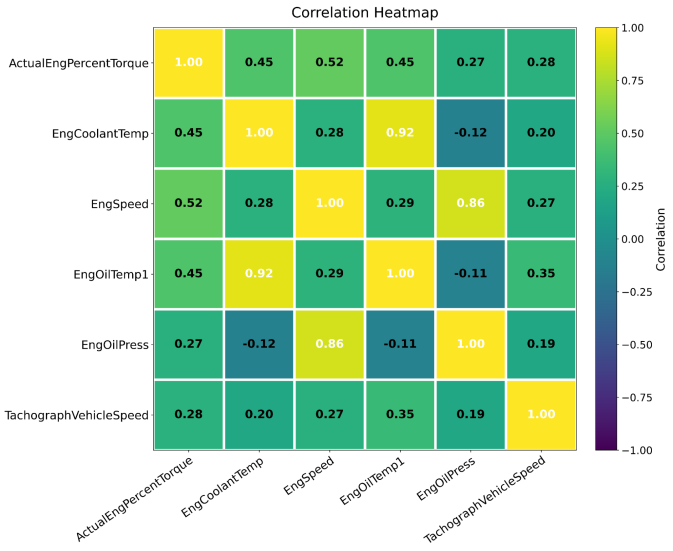

Figure 1. Correlation Heatmap of Vehicle Health Features.

Data preprocessing is a critical step in preparing the collected dataset for analysis. Preprocessing techniques such as standardization, null removal and data type conversion are applied to ensure usability for anomaly detection. Feature correlation is extracted by using a heatmap as in Figure 1. For instance, *EngCoolantTemp* and *EngOilTemp1* exhibit a strong positive correlation (0.92), indicating that engine load and thermal dynamics influence both parameters simultaneously. Similarly, *EngSpeed* and *EngOilPress* display a high correlation (0.86), reflecting the mechanical relationship between engine speed and oil pressure. Features with a correlation higher than 0.3 with any other feature is selected. The selected features are; actual engine torque percent (*ActualEngPercentTorque*), engine coolant temperature (*EngCoolantTemp*), engine speed (*EngSpeed*), engine oil temperature (*EngOilTemp1*), engine oil pressure (*EngOilPress*), tachograph vehicle speed (*TachographVehicleSpeed*). 43858 samples were obtained from 6 features in total by data preprocessing. The data is normalized through a standard scaler by setting $\sigma = 1$ and $\mu = 0$, for each feature. The preprocessed data is fed into the K-Means, Isolation Forest, Z-Score, GMM and DBSCAN algorithms to perform anomaly detection.

III. Anomaly Detection Algorithms

We implement five different algorithms to analyze the preprocessed dataset for anomaly detection: K-Means Clustering, Isolation Forest, Z-Score Analysis, DBSCAN and GMM Clustering.

K-Means clustering is initially utilized to obtain a basic set of clusters. The clustering results are analyzed using PCA, which reduces the high-dimensional data to two components for easy interpretation. Data points deviating farther than 2 standard deviations from the mean distance of each cluster are flagged as anomalies. This method's suitability lies in its ability to capture distinct operational patterns, such as idling or high-speed driving, and effectively highlight deviations from these norms.

Isolation Forest, a machine learning algorithm specifically designed for anomaly detection, is applied to the same dataset. It isolates anomalies by constructing random decision trees and measuring path lengths. Its ability to handle high-dimensional and diverse feature distributions makes it particularly effective for identifying unusual engine states or sensor malfunctions. The algorithm's parameters, such as contamination level, number of estimators, max samples and max features are optimized through grid search. The grid for the number of estimators consists of the values 100, 200, 500. The grid for the maximum samples consists of automatic selection, 50%, 75% of the dataset. The grid for the contamination factor consists of 1%, 5%, 10% of the data. The grid for the maximum feature at each tree consists of 100%, 50% of the features. Isolation Forest is utilized, along K-Means to detect anomalies in the vehicle sensor data by isolating outlier instances based on randomly constructed decision trees, making it suitable for identifying rare and irregular driving patterns or sensor behaviors.

Z-Score Analysis provides a statistical perspective by calculating the number of standard deviations each data point deviates from the mean. Z-Score anomalies are primarily observed in features with normal distributions, such as engine coolant temperature and oil pressure, to identify values outside typical operating ranges. A threshold of three standard deviations is used to flag extreme outliers. Such a threshold is chosen because approximately 99.7% of data points lie within 3 standard deviations from the mean.

GMM, a probabilistic clustering technique that represents each observation as arising from a mixture of multivariate Gaussian distributions with their own means, covariances and standard deviations, is applied to the vehicle dataset.

This flexibility in modeling both the central tendencies and the spread of correlated features (e.g., engine temperature, oil pressure, speed) makes GMM particularly effective for capturing overlapping driving modes and for identifying subtle deviations indicative of early-stage faults. The parameters for GMM is selected through grid search. The grid for the number of components consists of 2, 3, 4, 5, 6. The grid for the covariance matrix type consists of full, tied, diagonal and spherical. The best results are obtained with 2 components and a tied covariance matrix.

DBSCAN is a density-based clustering algorithm that groups together points in high-density regions while marking points in low-density regions as noise. In the context of vehicular data clustering, DBSCAN excels at discovering clusters of typical operational states (cruising, idling, or acceleration patterns) without requiring the number of clusters a priori. Through grid search, the best DBSCAN results are obtained with an epsilon value of 1.120833 and minimum sample value of 15. The grid for epsilon consists of 50 linearly spaced values running from 0.1 to 5.0. The grid for the minimum samples consists of the values 3, 5, 10, 15.

K-Means and GMM are utilized to capture the global structure of the dataset. DBSCAN and Isolation Forest spot local and high-dimensional irregularities; and Z-Score provides a transparent, per-signal sanity check. Their ensemble leverages diverse detection philosophies such as centroid distance, density isolation, distribution probability, tree-based partitioning, and statistical thresholding. Combining the strengths of each algorithm proves to be far more robust than using any single method.

IV. Performance Evaluation

The data is preprocessed and analyzed using Python and libraries such as Scikit-Learn, NumPy, SciPy, Matplotlib, Optuna, Seaborn and Pandas. The experiments are run in an NVIDIA Drive AGX Orin device. The analysis employed K-Means Clustering, Isolation Forest, Z-Score, DBSCAN and GMM methods to identify anomalies in connected truck telemetry data. Anomalies are defined as indicators of extreme aggression in vehicle driving behavior or internal vehicle mechanics that could lead to future vehicle failures. For instance, a data point with an *ActualEngPercentTorque* value of 94.0, a *EngCoolantTemp* value of 100.0, a *EngSpeed* value of 1386.0, a *EngOilTemp1* value of 119.65625, a *EngOilPress* value of 300.0, a *TachographVehicleSpeed* value of 38.210938 is flagged as an anomaly, due to having *ActualEngPercentTorque* value higher than the threshold of 90 and a *EngOilTemp1* value of 119.65625 higher than the threshold of 110.

A. Evaluation Metrics

To assess the effectiveness of the anomaly detection algorithms, several evaluation metrics are employed. These metrics are selected to provide insights into the clustering performance, agreement between methods, and the ability of each algorithm to identify meaningful anomalies within the dataset.

The performance of Z-Score Analysis and Isolation Forest is further validated by comparing their flagged anomalies against domain-specific thresholds for engine-related features. The threshold for *ActualEngPercentTorque* is set at 90%, similar to [19], which maps a John Deere 4.5 L engine's torque curve and identified that 90 % of rated torque corresponds to a high-load operating point. The threshold for the variable *EngCoolantTemp* is chosen to be 100 degrees Celsius. [20] proves that temperatures approaching or exceeding 100 degrees Celsius precipitate notable performance losses and component damage. *EngSpeed* as a threshold is set to 2500 revolutions per minute. In [21], the upper bound for a Caterpillar 3126B engine is found to be 2300 revolutions per minute, for emissions purposes. Since our work focuses on testing heavy duty vehicle load, we have set the threshold to be slightly higher than the 2300 rpm limit. *EngOilTemp1* is set to have a threshold of 110 degrees Celsius. [22] shows that a heavy-duty single-cylinder diesel engine achieved its lowest specific fuel consumption 83 degrees Celsius and 88 degrees Celsius. Their work argues that beyond 95 degrees Celsius, engine oil consumption begins to increase again. Industry-standard ageing tests show that key anti-wear additives degrade markedly once oil temperatures exceed 110 degrees Celsius [23]. *EngOilPress* is set at 600 kPa as a threshold. The value is set at such a level according to [24]. In the work, the authors have set their maximum measurement interval between 0-700 kPa for the engine oil pressure, where the majority of the points lie within 0-600 kPa. The points above 600 kPa are thus treated as anomalies in our work. *TachographVehicleSpeed* is set at 90 kmh. The threshold was chosen according to [25], where they demonstrate that carbon monoxide and soot spikes occurred in the interval of 100–130 kmh of vehicle speed. Since high carbon monoxide and soot generation of the system means erroneous operation, a speed value of 90 km/h serves as a good barrier for prevention. The silhouette score is used to evaluate the quality of the K-Means clustering. This metric measures the cohesion and separation of clusters by comparing the average distance of points within a cluster to the average distance to points in the nearest cluster [26]. A higher silhouette score indicates well-defined and distinct clusters.

B. Evaluation Setup

The Calinski-Harabasz Index is utilized as an internal validation metric to assess the clustering quality in the K-Means algorithm. This index evaluates the ratio of dispersion between groups to dispersion within groups, effectively measuring how well defined and compact clusters are [27]. A higher Calinski-Harabasz score indicates that the clusters are both tightly packed and well separated from each other. Furthermore, [28] highlights the index's robustness in evaluating clustering performance across varying datasets and its sensitivity to data separability, making it a reliable metric for assessing the clustering model's effectiveness. This metric complements the silhouette score by providing an additional perspective on the

effectiveness of the clustering model in distinguishing between normal and anomalous behaviors within the dataset.

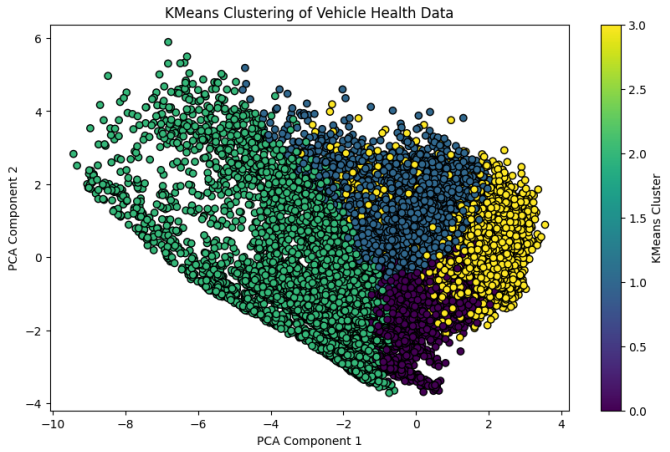

Figure 2. PCA with 2 Components Through K-Means.

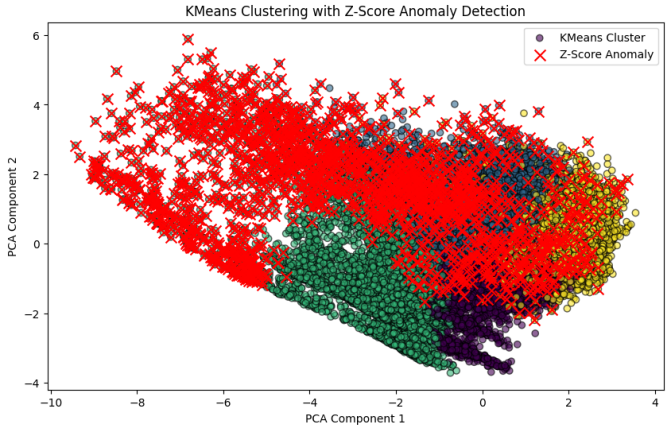

Figure 3. PCA Component Analysis Through K-Means and Z-Score Anomaly Detection.

During K-Means analysis, four clusters are initially utilized to capture major operational modes of the vehicle; idle, cruise, acceleration, deceleration. Each operational mode has its own level of engine load and thus needs its own analysis. Utilizing K-Means clustering and PCA, the clusters in Figure 2 are obtained. Each color in the cluster diagram denotes a unique mode of vehicle operation. Then, obtained clusters are relayed to the Z-Score algorithm, for further processing of anomalies and detection thereof. The Z-Score algorithm is then run with a Z-Score threshold value of 3. The resulting anomaly detection plot is given in Figure 3, where the cross-markers denote the anomalies. The Isolation Forest algorithm is then run with a contamination factor of 0.05 and a hundred estimators, to further detect anomalies or modify the existing anomalies.

Concurrently, the GMM is utilized. The pre-processed data is further processed for utilizing in GMM. Sorting, windowing and sampling operations are done on data before utilizing GMM. Sorting is done as per the timestamp of the data. For windowing, data is first resampled by splitting the data into contiguous, non-overlapping 60-second bins, for each feature. Then, the data is windowed to bins of separate statistic types, namely the mean, the standard deviation, the minimum, the maximum of each 60 second period, for each feature. Then, columns with an 80 percent of missing values or more are dropped. After windowing is completed, the data is imputed and scaled via a standard scaler. The GMM algorithm is run with a tied covariance matrix and 2 components. Similarly, after hyperparameter optimization, the DBSCAN algorithm is run with an epsilon value of 1.121, minimum sample value of 15.

C. Evaluation Results

In method evaluation, the silhouette score confirms the suitability of the chosen number of clusters, reflecting clear separations between operational patterns in the truck's telemetry data. On the other hand, the Calinski-Harabasz Index validates the clustering solution by confirming the distinctiveness of the operational patterns identified in the truck's telemetry data.

TABLE I.
COMPARISON OF CLUSTERING AND ANOMALY-DETECTION METHODS ON VEHICLE TELEMETRY DATA.

METHOD	SILHOUETTE	CALINSKI–HARABASZ
K-Means	0.33	**19281.7**
Isolation Forest	0.473	671
GMM	0.513	6379.11
DBSCAN	0.297	5.916
Z-Score	**0.569**	7777.99

Z-Score analysis resulted in the highest silhouette score of 0.569 and a strong Calinski–Harabasz Index of 7777.99, reflecting its straightforward yet precise anomaly detection. GMM followed with a silhouette score of 0.513 and an index of 6379.11, demonstrating robustness in modeling both central tendencies and spread. Isolation Forest achieved a silhouette of 0.473 but a low Calinski–Harabasz Index of 671, indicating its specialization in outlier isolation over cluster cohesion. K-Means showed moderate performance with a silhouette score of 0.33, Calinski-Harabasz index 19281.7 but was less effective at capturing anomalies. DBSCAN's density-based grouping yielded a low silhouette score of 0.297 and an index of 5.916, highlighting its difficulty in forming cohesive clusters for this dataset.

V. Conclusions and Future Work

This paper presents a unified benchmarking framework for anomaly detection on engine-related connected-truck onboard telemetry using K-Means, Isolation Forest, Z-Score analysis, GMM, and DBSCAN. The results indicate that Z-Score achieved the best overall performance on the studied dataset, providing a transparent and effective per-signal sanity check under domain-informed thresholds. K-Means was useful for separating major operational modes, while Isolation Forest complemented the analysis by isolating rare, high-dimensional outliers. Although DBSCAN can explicitly label low-density samples as noise and discover non-spherical structures, its

single-density assumption may lead to fragmented clusters when the data exhibit varying densities. In contrast, GMM provides a probabilistic soft-assignment model that can represent overlapping operating regimes and correlated feature structures, which is valuable for capturing subtle deviations.

Future work will focus on (i) validating the proposed approach on longer time horizons and larger fleets to assess robustness under seasonal and operational variability, (ii) extending the pipeline toward online/streaming anomaly detection to enable real-time edge deployment, and (iii) incorporating weak supervision from maintenance logs or diagnostic trouble codes to calibrate thresholds, reduce false positives, and improve interpretability. In addition, exploring temporal modeling and change-point detection on multivariate telemetry may further enhance early-fault sensitivity beyond point-wise outlier scoring.

REFERENCES

[1] A. Giannoulidis, K. Karras, G. Giannakopoulos, and V. Vassalos, "Exploring unsupervised anomaly detection for vehicle predictive maintenance with partial information," in *Proceedings of the 27th International Conference on Extending Database Technology (EDBT)*, 2024, pp. 753–761, paestum, Italy. [Online]. Available: https://openproceedings.org/2024/conf/edbt/paper-200.pdf

[2] I. J. Chukwudi *et al.*, "An ensemble deep learning model for vehicular engine health prediction," *IEEE Access*, vol. 12, pp. 63 433–63 451, 2024.

[3] O. Gheibi, D. Weyns, and F. Quin, "Applying machine learning in self-adaptive systems: A systematic literature review," *ACM Transactions on Autonomous and Adaptive Systems*, vol. 15, no. 3, pp. 1–37, 2021. [Online]. Available: https://dl.acm.org/doi/10.1145/3469440

[4] J. Selech and K. Przystupa, "Reliability analysis for unrepairable automotive components," *Applied Sciences*, vol. 11, no. 22, p. 10863, 2021.

[5] F. Arena, M. Collotta, L. Luca, M. Ruggieri, and F. G. Termine, "Predictive maintenance in the automotive sector: A literature review," *Mathematical and Computational Applications*, vol. 27, no. 1, p. 2, 2022. [Online]. Available: https://www.mdpi.com/2297-8747/27/1/2

[6] Y. Cherdo, B. Miramond, A. Pegatoquet, and A. Vallauri, "Unsupervised anomaly detection for cars can sensors time series using small recurrent and convolutional neural networks," *Sensors*, vol. 23, no. 11, p. 5013, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/11/5013

[7] R. Domingues, M. Filippone, P. Michiardi, and J. Zouaoui, "A comparative evaluation of outlier detection algorithms: Experiments and analyses," *Pattern Recognition*, vol. 74, pp. 406–421, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320317303916

[8] K. Demestichas, T. Alexakis, N. Peppes, and E. Adamopoulou, "Comparative analysis of machine learning-based approaches for anomaly detection in vehicular data," *Vehicles*, vol. 3, pp. 171–186, 2021.

[9] W. Jia and C. Yang, "A framework for multi-variate time series anomaly detection based on behavioral patterns in electric vehicles," in *2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, Dali, China, 2022, pp. 20–24.

[10] C. Derse *et al.*, "An anomaly detection study on automotive sensor data time series for vehicle applications," in *2021 Sixteenth International Conference on Ecological Vehicles and Renewable Energies (EVER)*, Monte-Carlo, Monaco, 2021, pp. 1–5.

[11] C. H. M. Chandralekha, H. M. Chandrashekar, P. S. Nijesh, S. P. S. Pai, and M. K. Ghosh, "Anomaly detection in recorded can log using dbscan and lstm autoencoder," in *2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT)*. Bangalore, India: IEEE, 2022, pp. 1–7.

[12] F. Fenzl, R. Rieke, Y. Chevalier, A. Dominik, and I. Kotenko, "Continuous fields: Enhanced in-vehicle anomaly detection using machine learning models," *Simulation Modelling Practice and Theory*, vol. 105, p. 102143, 2020.

[13] S. Park and J.-Y. Choi, "Hierarchical anomaly detection model for in-vehicle networks using machine learning algorithms," *Sensors*, vol. 20, no. 14, p. 3934, Jul 2020.

[14] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 Eighth IEEE International Conference on Data Mining*, 2008, pp. 413–422.

[15] Z.-G. Zhou and P. Tang, "Continuous anomaly detection in satellite image time series based on z-scores of season-trend model residuals," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 3410–3413.

[16] H. Davoudi, B. S. Ghahfarokhi, N. Moghim, and S. Shetty, "Dbscan-based vehicle clustering and uav placement for noma-based resource management in cellular v2x communications," *arXiv preprint arXiv:2504.21656*, 2025. [Online]. Available: https://arxiv.org/abs/2504.21656

[17] B. Wang, W. Li, and Z. H. Khattak, "Anomaly detection in connected and autonomous vehicle trajectories using lstm autoencoder and gaussian mixture model," *Electronics*, vol. 13, no. 7, p. 1251, 2024. [Online]. Available: https://www.mdpi.com/2079-9292/13/7/1251

[18] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, 2020. [Online]. Available: https://www.mdpi.com/2079-9292/9/8/1295

[19] R. A. Rohrer, J. D. Luck, S. K. Pitla, and R. Hoy, "Evaluation of the accuracy of machine reported can data for engine torque and speed," *Transactions of the ASABE*, vol. 61, no. 5, pp. 1547–1557, 2018.

[20] J. Adler and T. Bandhauer, "Performance of a diesel engine at high coolant temperatures," *Journal of Energy Resources Technology*, vol. 139, no. 6, p. 062203, 2017.

[21] T. Huai *et al.*, "Analysis of heavy-duty diesel truck activity and emissions data," *Atmospheric Environment*, vol. 40, no. 13, pp. 2333–2344, 2006.

[22] S. Pukalskas, A. Pikūnas, and G. Bureika, "The influence of diesel engine's thermal condition on fuel consumption and smoke of exhaust gases," *Journal of KONES: Powertrain and Transport*, vol. 16, no. 3, pp. 319–324, 2009.

[23] ASTM International, "Astm d7528-09: Standard test method for bench oxidation of engine oils by robo apparatus," ASTM International, West Conshohocken, PA, Standard D7528-09, 2009.

[24] A. Grzesiek, R. Zimroz, P. Śliwiński, N. Gomolla, and A. Wyłomańska, "A method for structure breaking point detection in engine oil pressure data," *Energies*, vol. 14, no. 17, p. 5496, 2021.

[25] J. Gao, H. Chen, K. Dave, J. Chen, and D. Jia, "Fuel economy and exhaust emissions of a diesel vehicle under real traffic conditions," *Energy Science & Engineering*, vol. 8, no. 5, pp. 1781–1792, 2020. [Online]. Available: https://doi.org/10.1002/ese3.632

[26] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," in *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, 2020, pp. 747–748.

[27] S. Łukasik, P. A. Kowalski, M. Charytanowicz, and P. Kulczycki, "Clustering using flower pollination algorithm and calinski-harabasz index," in *2016 IEEE Congress on Evolutionary Computation (CEC)*, 2016, pp. 2724–2728.

[28] X. Wang and Y. Xu, "An improved index for clustering validation based on silhouette index and calinski-harabasz index," *IOP Conference Series: Materials Science and Engineering*, vol. 569, no. 5, p. 052024, jul 2019.

A Comparison of Public Electric Vehicle Charging Access across the United States

Monika M. Wahi
Research and Data Lab
New Delhi, India
Email: mwahi@RADL.online

Dinesh Chacko
Institute of Electrical and Electronic Engineers
Maidenhead, United Kingdom
Email: dinesh.chacko@ieee.org

***Abstract*—While the rate of new electric vehicle registrations was only 10% in 2023 in the United States, electric vehicle users still need access to public charging points. This study aimed to characterize the differences in public charging point access across the United States, and to estimate if low access was associated with state-level rurality and low socioeconomic conditions. Based on data available from the United States census and Department of Energy, state-level public charging point access was defined two ways: As public charging points per 10,000 residents, and as a ratio of registered electric vehicles per public charging point. Descriptive, geographical, correlation and regression analysis was conducted to explore associations between socioeconomic variables and public charging point access between states, and to compare access between states. When comparing all states and the District of Columbia (n = 51), we found that public charging points per 10,000 residents ranged from 0.9 to 9.9 (median 2.4, interquartile range 1.7 to 3.4), and electric vehicles per public charging point ranged from 9.2 to 76.5 (median 30.4, interquartile range 19.3 to 39.7). In regression analysis, median state income was statistically significantly negatively associated with public charging points per 10,000 residents, and percentage with college degrees was strongly statistically significantly positively associated with this rate. No socioeconomic variables were statistically significantly associated with electric vehicles per public charging point.**

***Keywords*—Electric vehicles; public charging points; United States; socioeconomic status; public policy.**

I. INTRODUCTION

Electric Vehicle (EV) ownership is low in the United States (US) relative to other countries; while the rate of new EV registrations in the US was 10% (representing 1.4 million) in 2023, it was just under 25% in Europe and about 60% in China in the same year [1]. EVs must be charged, and in 2021, the National Renewable Energy Laboratory (NREL) estimated that between 78% and 98% of US EV owners had access to residential charging [2]. It has been estimated in the US in 2023 that 50% to 80% EV charging sessions occurred at the residence, although less than 50% of households have parking areas available where residential EV charging could take place [3]. For example, residents of multi-unit dwellings in the US have specific challenges to setting up residential EV charging [2][3].

Setting aside the obstacles of establishing EV charging at certain residence types, another important downside to residential EV charging is that it strains the energy grid [4]. An additional consideration is that all EV charging cannot take place residentially if EV users want to take long trips that necessitate access to chargers available in the community [5][6]. To account for this in the US, the Biden administration set a goal of creating a national network with 500,000 EV public Charging Points (CPs), and a law was passed providing $7.5 billion to support this expansion [5][7]. Under a scenario where 50% of all new vehicle sales are Zero-Emission Vehicles (ZEVs) by 2030 (to be in line with federal targets), it is estimated that the US will need to have established approximately 1.2 million public EV CPs and 28 million private chargers by that year [8][9].

Although many other countries have also invested in their public EV charging infrastructures, disparities have been identified with respect to public CP accessibility [10]. Studies in the US and other countries show that those who purchase EVs are more likely to have higher income (largely due to the high EV initial purchase price) and educational attainment [11–16] . One study has shown that the availability of public CPs in the US is correlated with the density of EV ownership, which illustrates one aspect of such disparities [17]. This issue has been termed the “chicken and egg” problem, whereby public CPs are set up where EV ownership is higher, and this incentivizes more EV ownership in the region where public CPs are established [18]. Nevertheless, other studies in the US have found that the availability of public CPs is not correlated with population density, and is instead associated with higher income, older age, larger percentage those identifying as white race, and proximity to highways [13][14][18][19]. Another US study found that disadvantaged communities (defined as those exposed to a disproportionate share of environmental, health, and climate-related burdens) have 64% fewer public CPs compared to other communities [20]. In the US and in other countries, the availability of public CPs has been found to be lower in rural areas [21][22].

Public CP access is generally compared between regions using either CPs per population, or CPs per EV [6][23][24]. Zema et al. [24] developed a top ten list of European countries with the highest density of public CPs per population, and the lowest rates of EVs per public CP. Their analysis found that densities of public CPs per 10,000 population for the top ten European countries ranged from 10.4 to 52.5, and the densities of EVs per public CP for the top ten European countries ranged from 0.6 to 3.0 [24].

While several research reports identify factors associated with public CP access disparities in the US, their approaches typically do not look at the differences between states [10][19–21] . For a few examples, one study of inequity in public CP access in the US examines this phenomenon at the census track level [21]; others consider attributes of US

regions in proximity to public CPs without regard to political boundaries (such as census tracts, counties, or states) [10][19]; another compares public CP access in US communities with particular attributes [20]; and others investigate public CP access within specific locations, such as King County in Washington state [25], New York City (NYC) [26], and California [27]. These approaches, while legitimate, cannot take into account the potential impact of state-level policies supporting EV ownership and the development of a public EV infrastructure [26][27]. California has passed laws and funded projects aimed at increasing EV ownership and public CP access, and projects to increase access to public CPs are also underway in NYC, Kansas City in Missouri, and in the states of Washington, New Jersey, and Ohio [13][26][27].

Additionally, socioeconomic conditions differ between states which can impact disparities between states in terms of the public CP infrastructure [13][28]. States also differ in their proportions of non-white residents, low income households, and rural locations, attributes which have been shown to be associated with increased energy burden [28]. Also, states have different levels of housing types (such as multi-unit dwellings) that can impact the ability of residents to engage in EV charging at home [13]. Therefore, the aim of this investigation was to characterize the differences in public CP access across the states in the US, and to estimate if low access was associated with state-level rurality and low socioeconomic conditions.

In this paper, Section II presents our methodology, including the data sources we used and our analytic approach. Section III provides our results, including a descriptive analysis, followed by correlations, Analysis of Variance (ANOVA), and regression analysis. Section IV presents a discussion, including the strengths and limitations of the analysis, and the paper concludes with Section V, which provides a conclusion and discusses future work.

II. Methodology

A. Data Sources

For this analysis, each state in the US (including Washington DC) was considered an experimental unit (n = 51). State-level rurality was operationalized as non-urban areas and was estimated from the gazetteer files provided by the US census as a proportion. In the proportion, the numerator was estimated by subtracting Urban Square Miles (SQMI, represented by SQMI for metropolitan areas) from total state SQMI [29]. The denominator was estimated as total state SQMI. Where metropolitan areas crossed multiple states, they were divided equally among each state. State-level socioeconomic conditions were operationalized as median state income (median earnings in the past 12 months in 2024 inflation-adjusted dollars for the population aged 25 years and over with earnings) and proportion of residents age 25 years and over with bachelor's degree or higher 2024. These data were obtained online for each state from the US census [30].

Public CP access was operationalized in two ways: As number of public CPs per 10,000 state population (*CPPrev*), and as a ratio of number of EVs the state divided by number of CPs in the state (*EVPerCP*). Number of EVs per state registered in 2024 and number of CPs per state were obtained from the Alternative Fuels Data Center (AFDC) provided by the US Department of Energy; CP data were filtered to only include EV CPs that were available as of September 16, 2025 [31]. State population was estimated using the US census [30]. States were also analyzed in terms of divisions and regions (see Table 1).

B. Analytic Approach

First, a descriptive analysis was performed on each state individually, including mapping, then aggregated by division and region. Next, correlations were explored between predictors and public CP access measures (*CPPrev* and *EVPerCP*). ANOVAs were conducted with region and division as predictors and public CP access variables (*CPPrev* and *EVPerCP*) as dependent variables.

Finally, a linear regression model was developed for each public CP access measure as the dependent variable including the independent variables of proportion non-urban, median income, proportion with a college degree, and division as indicator variables (compared to the East South Central division, see Table 1). For the regression model predicting *CPPrev* (Model A), 1,000s of registered EVs in 2024 was included as an independent variable, and for the regression model predicting *EVPerCP* (Model B), 10,000 of population was included as an independent variable. For all analyses with statistical tests, α was set at 0.05. For regression models, variables were retained if they met or approached statistical significance. Statistical analysis was done using R GUI [32] and maps were developed in Python.

III. Results

A. Descriptive Analysis

Descriptive statistics were developed to characterize each state, including adult population, proportion of non-urban area, median income, and proportion of the population with college degrees (data not shown). The results were that the median population of all the states was 3,531,346 (Interquartile Range [IQR] 1,471,384 to 6,148,676), the median proportion non-urban was 0.98 (IQR 0.94 to 0.99), the median annual income was $51,472 (IQR $50,113 to $57,208), and the median proportion with college degrees was 0.36 (IQR 0.32 to 0.39). Descriptive statistics regarding EVs and CPs were also developed for states (data not shown). The median number of CPs per state was 714 (IQR 375 to 1,899), and the median *CPPrev* was 2.4 (IQR 1.7 to 3.4, range 0.9 to 9.9). The median number of EVs per state was 25,565 (IQR 8,108 to 71,152) and the median *EVPerCP* was 30.4 (IQR 19.3 to 39.7, range 9.2 to 76.5). These results are summarized by division and region in Table 1.

As shown in Table 1, *CPPrev* ranged from 1.6 to 6.1, and *EVPerCP* ranged from 18 to 61. Figure 1 shows the number of EVs and public CPs per state. In Figure 1, the category boundaries follow close to the quartile estimates to facilitate

TABLE I
DESCRIPTIVE STATISTICS ABOUT ELECTRIC VEHICLES AND CHARGING POINTS BY DIVISION AND REGION

Division*	Number of States	Number of Public Charging Points	Charging Points per 10,000 Populations	Number of Registered Electric Vehicles 2024	Electric Vehicles per Public Charging Point
ENC	5	7,290	1.9	251,294	34
ESC	4	2,447	1.6	61,475	25
MA	3	9,111	2.7	336,157	37
MTN	8	6,859	3.4	291,764	43
NE	6	7,611	6.1	136,775	18
PAC	5	24,572	5.8	1,501,370	61
SA	9	13,931	2.5	614,617	44
WNC	7	4,112	2.4	93,767	23
WSC	4	5,320	1.6	268,226	50
Region**					
MW	12	11,402	2.1	345,061	30
NE	9	16,722	3.6	472,932	28
S	17	21,698	2.1	944,318	44
W	13	31,431	5.0	1,793,134	57

Note: * ENC = East North Central, ESC = East South Central, MA = Middle Atlantic, MTN = Mountain, NE = New England, PAC = Pacific, SA = South Atlantic, WNC = West North Central, WSC = West South Central. ** MW = Midwest (includes WNC and ENC), NE = Northeast (includes MA and NE), S = South (includes WSC, ESC and SA), W = West (includes MTN and PAC).

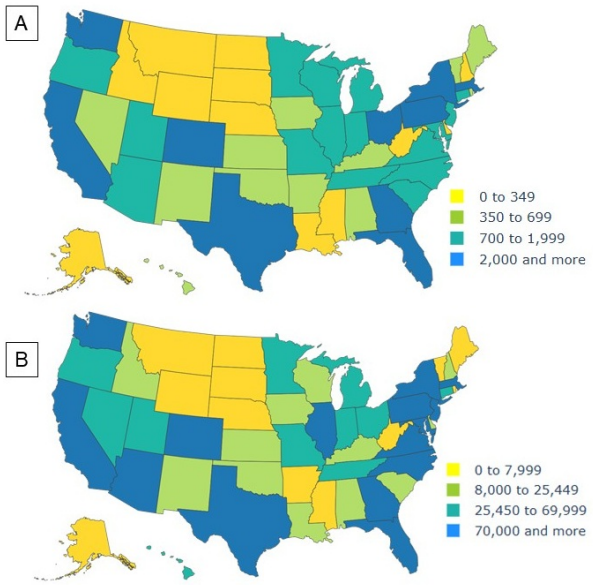

Figure 1. Number of Electric Vehicles and Public Charging Points per State. A: Number of public charging points. B: Number of electric vehicles.

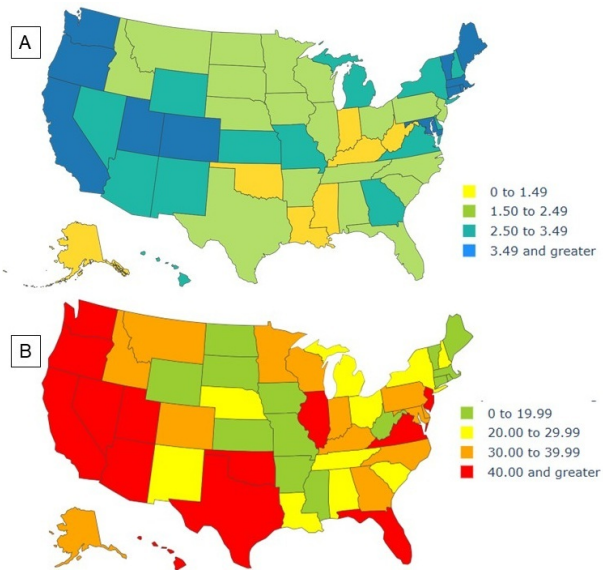

Figure 2. Public Electric Vehicle Charging Points by 10,000 Population and by Number of Electric Vehicles per State A: Public electric vehicle charging points per 10,000 population. B: Electric vehicles per public charging point.

a fair comparison. As shown in Figure 1, there are slight differences in the density of EVs compared to the density of public CPs per state across states. For example, several states in the South Atlantic (SA), East North Central (ENC) and Mountain (MTN) divisions appear to have a higher density of EVs compared to public CPs. Figure 2 visualizes *CPPrev* and *EVPerCP* by state. As with Figure 1, in Figure 2, the category boundaries roughly follow quartiles to facilitate a fair comparison. In Figure 2, although states in the West region have high *CPPrev*, they also have a high *EVPerCP*. States with low *CPPrev* and high *EVPerCP* include Illinois, New Jersey, Florida, Oklahoma and Texas, suggesting EV owners have the most limited access to public CPs in those states. Several states in the Northeast region have high *CPPrev* but low *EVPerCP*, suggesting that EV adoption by the population is outpaced by the public CP charging infrastructure.

B. Correlation and Analysis of Variance

In correlation analysis, adult population was strongly positively statistically significantly correlated with number of registered EVs in the state (r = 0.8054, $p < 0.0001$) and number of public CPs in the state (r = 0.8448, $p < 0.0001$), and number of registered EVs was even more strongly positively statistically significantly correlated with number of public CPs (r = 0.9747, $p < 0.0001$). In terms of access measures, with respect to *CPPrev*, a strong positive statistically significant correlation was seen with proportion with college degrees (r = 0.7035, $p < 0.0001$), and a moderate positive statistically significant correlation was found with state median income (r = 0.5715, $p < 0.0001$). A weak but statistically significant negative correlation was seen between *CPPrev* and proportion of non-urban land (r = -0.3631, p = 0.0088). In terms of *EVPerCP*, there was a moderate, positive statistically significant correlation with state adult population (r = 0.4847, p = 0.0003). All other correlations were not statistically significant.

For ANOVA results, with respect to *CPPrev*, both division (p = 0.0001) and region (p = 0.0007) were statistically significant, and similar results were found for *EVPerCP* (division p = 0.0054, region p = 0.0399).

C. Regression Analysis

Both Model A (with *CPPrev* as the dependent variable) and Model B (with *EVPerCP* as the dependent variable) were developed. To develop a final equation for each model, for Model A, only the following covariates were retained in the model: 1,000 registered electric vehicles 2024, median income, proportion with college degrees, and the indicator variable for NE. For the final Model B equation, the following covariates were retained: 10,000 population and the indicator variable for PAC. The results were:

Model A

$$\hat{y} = -1.1959 + 0.0026_{\text{EV}} - 0.0001_{\text{MI}} + 23.4885_{\text{PCD}} + 2.3771_{\text{NE}}$$

Model B

$$\hat{y} = 24.8170 + 0.0121_{\text{POP}} + 16.7633_{\text{PAC}}$$

where *EV* is number of EVs in the state, *MI* is median income, *PCD* is proportion of residents with college degrees, *NE* is New England division, *POP* is adult population, and *PAC* is Pacific Division. The adjusted r^2 for final Model A was 0.7126, and for final Model B, the adjusted r^2 was 0.2935. These measures indicate that while the covariates in Model A explained about 71% of variance in *CPPrev*, the covariates in Model B only explained about 29% of the variance in *EVPerCP*, indicating that the model fit for Model B was far inferior to the model fit for Model A. While the higher r^2 for Model A indicates strong in-sample explanatory power, this does not by itself guarantee predictive performance on new data, which is sufficient for the current analysis which is not intended to be predictive.

In Model A, which predicts *CPPrev*, 1,000 registered EVs was statistically significant (slope = 0.0026, p = 0.0016), as was proportion with college degrees (slope = 23.4885, $p <$ 0.0001) and the NE division (slope = 2.3771, $p < 0.0001$). All other covariates in the model were not statistically significant. However, median income, which had a negative association with *CPPrev* in the model, approaches statistical significance (slope = -0.0001, p = 0.0627). As those with college degrees typically have a higher income than those without, the contradiction between the positive slope for proportion with college degrees and the negative slope for median income may reflect collinearity between the two covariates in the model. All statistically significant slopes were positive, indicating a positive association with *CPPrev*, and the magnitude of the slope for proportion with college degrees was much higher than the others, indicating that controlling for other factors, this covariate has a very strong positive association with *CPPrev*.

The results from Model B, which predicts *EVPerCP*, were quite different from Model A. The slope for the covariate for 10,000 population was positive and statistically significant (slope = 0.0121, p = 0.0009), as was the slope for Pacific division (slope = 16.7633 p = 0.0168). As was seen in Model A for the slope for proportion with college degrees, the magnitude of the slope for the Pacific division in Model B was very high. None of the other covariates are statistically significant in the model.

IV. Discussion

Our analysis showed that there were wide differences in public CP access across the different states in the US. In final regression models, neither measure of public CP access we used was associated with state-level rurality. However, the access measure of public CPs per 10,000 population was slightly negatively associated with median income and strongly positively associated with the proportion of population holding a college degree in the state. While the correlation analysis revealed a moderate positive statistically significant correlation between public CPs per 10,000 population with state median income, this association became slightly negative when included in the full regression model in the presence of other covariates. The access measure of EVs per public CPs in the state was not associated with any socio-economic variables in final models. In final models, it was found that state division membership mattered, in that for public CPs per 10,000 population, the NE division was strongly independently positively associated, and for EVs per public CP, the PAC division was strongly independently positively associated.

Several important observations can be made from this analysis. First, public CPs per 10,000 residents in each state ranged from 0.9 to 9.9. With respect to that metric, even the top state is lower than the lowest of the top ten European countries, which is Denmark at 10.4 [24]. With respect to EVs per public CP, the range for states was 9.2 to 76.5, which is much higher than the least accessible of the top ten European countries, which is Hungary at 3.0 [24]. This shows that while there is much variation state to state on these access metrics, overall, compared to European countries, public CPs are generally less accessible in the US.

Next, socioeconomic indicators in this analysis were associated with public CPs per 10,000 residents in each state, but not with EVs per public CP. Median income had a slight negative association, showing that higher median income is associated with fewer public CPs per 10,000 residents. However, proportion with college degrees in the state had a strong, positive association, suggesting that how educated the population is strongly influences the number of public CPs. As described earlier, EV owners tend to be more highly educated [11–16] , yet proportion of college degrees was not correlated with number of EVs in the state. Therefore, the persistence of college degrees being associated with higher rates of public CPs per population in the regression analysis is likely a result of a more educated populace selecting leaders that prioritize CP access as part of public policy. Although having a college degree is generally associated with enjoying a higher income, states with depressed economies generally have a lower median income, regardless of the level of education of their population.

That the NE division was strongly positively associated with CPs per population, and that the PAC division was

strongly positively associated with EVs per CP also is likely a reflection of public policy initiatives. With respect to the NE division, public policy has emphasized expanding the public CP infrastructure, and programs to expand public CP access have been implemented in Rhode Island, Massachusetts, and New Hampshire [33–35]. However, the states in PAC had significantly higher EVs per public CP, which suggests that CP access is much poorer in that division. This is likely due to an emphasis in public policy toward EV adoption without a comparable emphasis on expanding the public CP infrastructure to accommodate the increasing number of EV users. The California ZEV program, which was started in 1990, aims to reduce Greenhouse Gas (GHG) vehicle emissions through establishing vehicle standards, and other PAC states such as Oregon and Washington (as well as states in other divisions) have developed their own programs based on ZEV [36][37]. Because the focus in the PAC division states has been on EV adoption rather than public EV charging infrastructure, it appears that EV ownership has now outpaced the development of public CP access. While many studies examine a lack of equity in EV adoption within California and other PAC states [25][27][38][39], comparing access to the public CP infrastructure between these states and states in other divisions has not been a focus of research (although it has been acknowledged that there is a tradeoff between subsidizing EV adoption and investing in a public CP infrastructure in California specifically [39].

This analysis has both strengths and limitations. The main strength is that it provides an evidence-based descriptive comparison between states with respect to public CP access. While this is a strength, it limits the granularity of the analysis; as an example, the center of a large city may not have a heightened demand for public CP access because of the availability of public transportation, while an area in the city with a lower population density may have a higher need for public CP access. Other limitations include using only estimated measures of public CP access and socioeconomic status, as well as having sample size limited to the number of states in the analysis (which is likely why other research on public CP access in the US tends to use smaller geographic regions for comparisons [10][19][21]. We acknowledge that our research did not find an association between rurality and charger access at the state level, which contradicts prior studies that find lower charger density in rural areas [5][10][21][33]. The discrepancy likely stems from our use of broad state level rurality percentages rather than more precise measures (such as rural tract counts, or distance to nearest charger). As a result, our analysis may underestimate the true rural access gap. Further, the regression analysis of EVs per public CP (Model B) showed a relatively poor fit, indicating that substantial unexplained variability remains, potentially due to omitted factors, measurement noise, or nonlinear relationships not captured by the linear specification.

V. Conclusion and Future Work

Our analysis aimed to characterize the differences in access to public CPs between states in the US, as well as seek evidence as to whether low access was associated with attributes correlated with low public CP access identified in existing analyses, namely state-level rurality and low socioeconomic conditions. We did this by utilizing public data to conduct a state-by-state comparison, and by using regression to estimate whether rurality and low socioeconomic conditions were associated with low public CP access while adjusting for other variables. While a state to state comparison did not find an association between rurality and public CP access, number of public CPs per 10,000 population was found to be associated with socio-economic variables. Further, states in the NE division were found to have significantly better public CP access, while states in the PAC division were found to have significantly worse public CP access (compared to other divisions). Overall, public CP access in states was much worse when compared to access in European countries.

We acknowledge that the current analysis reflects the status of EVs and CPs in the US in 2024 only. Given the rapid growth of EV adoption and public CP development, our findings may become outdated quickly. For future research, we intend to track EV ownership and public CP access longitudinally, which will facilitate observing the impacts of policy changes and market dynamics in the US over time. Also, while this analysis only explored the socioeconomic variables of income and education, future research will consider other equity dimensions such as race/ethnicity and age when studying which subpopulations are most affected by public CP scarcity. Future research could also consider areas at a higher level of granularity, such as regions within a particular metropolitan area.

While public CP access is likely influenced by many factors, one of the strongest influences is public policy. States with many registered EVs but low levels of public CP access should now turn their attention to developing public initiatives aimed at increasing number of public CPs, as well as increasing access to them through lowering charging costs or other incentives [25][27][33–39] . As EV adoption continues on an upward trajectory in the US, it will be necessary for federal and state policies to prioritize ensuring public CP access to keep pace with this increase. Future research should investigate the most efficient methods for expanding public CP access in the US, especially among states where EV ownership is already high.

References

[1] International Energy Agency, “Global EV outlook 2024: Moving toward increased affordability,” International Energy Agency, Tech. Rep., 2024. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2024, retrieved: August, 2025.

[2] Y. Ge, C. Simeone, A. Duvall, and E. Wood, "There's no place like home: Residential parking, electrical access, and implications for the future of electric vehicle charging infrastructure," National Renewable Energy Laboratory, Tech. Rep. NREL/TP-5400-81065, 2021. [Online]. Available: https://www.osti.gov/biblio/1825510, retrieved: October, 2025.

[3] X. Cheng and E. Kontou, "Estimating the electric vehicle charging demand of multi-unit dwelling residents in the United States," *Environ. Res.: Infrastruct. Sustain.*, vol. 3, no. 2, pp. 1–15, 2023.

[4] S. Powell, G. V. Cezar, L. Min, I. M. L. Azevedo, and R. Rajagopal, "Charging infrastructure access and operation to reduce the grid impacts of deep electric vehicle adoption," *Nat. Energy*, vol. 7, no. 10, pp. 932–945, 2022.

[5] L. Hanig et al., "Finding gaps in the national electric vehicle charging station coverage of the United States," *Nat. Commun.*, vol. 16, no. 1, pp. 1–13, 2025.

[6] F. Schulz and J. Rode, "Public charging infrastructure and electric vehicles in Norway," *Energy Policy*, vol. 160, pp. 1–31, 2022.

[7] B. Borlaug, F. Yang, E. Pritchard, E. Wood, and J. Gonder, "Public electric vehicle charging station utilization in the United States," *Transp. Res. Part D: Transp. Environ.*, vol. 114, pp. 1–34, 2023.

[8] P. Kampshoff, A. Kumar, S. Peloquin, and S. Sahdev, "Building the electric-vehicle charging infrastructure America needs," McKinsey Center for Future Mobility, Tech. Rep., 2022. [Online]. Available: https://www.mckinsey.com/industries/public-sector/our-insights/building-the-electric-vehicle-charging-infrastructure-america-needs, retrieved: November, 2025.

[9] S. Russo, B. Spiller, and R. Wilwerding, "Equity in electric vehicle charging infrastructure," Resources for the Future, Tech. Rep. Working Paper 24–14, 2024. [Online]. Available: https://www.rff.org/publications/working-papers/equity-in-electric-vehicle-charging-infrastructure/, retrieved: January, 2026.

[10] J. Lou, X. Shen, D. A. Niemeier, and N. Hultman, "Income and racial disparity in household publicly available electric vehicle infrastructure accessibility," *Nat. Commun.*, vol. 15, no. 1, pp. 1–12, 2024.

[11] Z. Dai, M. O. Rodgers, and R. Guensler, "Hybrid EV and pure BEV owners: A comparative analysis of household demographics, travel behavior, and energy use," National Center for Sustainable Transportation, Tech. Rep. NCST-GT-RR-23-20, 2023. [Online]. Available: https://ncst.ucdavis.edu/research-product/hybrid-ev-and-pure-bev-owners-comparative-analysis-household-demographics-travel, retrieved: September, 2025.

[12] J. Kennedy, K. James, and B. Hampson, "Public opinion and understanding of the impact of electric vehicles: A UK experience," *Adv. Soc. Sci. Res. J.*, vol. 10, no. 10, pp. 248–277, 2023.

[13] D.-Y. Lee, M. H. McDermott, B. K. Sovacool, and R. Isaac, "Toward just and equitable mobility: Socioeconomic and perceptual barriers for electric vehicles and charging infrastructure in the United States," *Energy Clim. Chang.*, vol. 5, pp. 1–19, 2024.

[14] D.-Y. Lee et al., "Does electric mobility display racial or income disparities? Quantifying inequality in the distribution of electric vehicle adoption and charging infrastructure in the United States," *Appl. Energy*, vol. 378, pp. 1–17, 2025.

[15] W. S. Loh and R. B. Noland, "Concerns expressed by used electric vehicle owners based on surveying social media," *Transp. Res. Part D: Transp. Environ.*, vol. 128, pp. 1–13, 2024.

[16] R. Tao, X. Yang, F. Hao, and P. Chen, "Demographic disparity and influences in electric vehicle adoption: A Florida case study," *Transp. Res. Part D: Transp. Environ.*, vol. 136, 2024.

[17] L. V. White, A. L. Carrel, W. Shi, and N. D. Sintov, "Why are charging stations associated with electric vehicle adoption? Untangling effects in three United States metropolitan areas," *Energy Res. Soc. Sci.*, vol. 89, 2022.

[18] E. Hopkins, D. Potoglou, S. Orford, and L. Cipcigan, "Can the equitable roll out of electric vehicle charging infrastructure be achieved?" *Renew. Sustain. Energy Rev.*, vol. 182, pp. 1–13, 2023.

[19] A. Ermagun and J. Tian, "Charging into inequality: A national study of social, economic, and environment correlates of electric vehicle charging stations," *Energy Res. Soc. Sci.*, vol. 115, 2024.

[20] Q. Yu et al., "Equity and reliability of public electric vehicle charging stations in the United States," *Nat. Commun.*, vol. 16, no. 1, pp. 1–13, 2025.

[21] G. J. Carlton and S. Sultana, "Electric vehicle charging equity and accessibility: A comprehensive United States policy analysis," *Transp. Res. Part D: Transp. Environ.*, vol. 129, 2024.

[22] T. R. McKinney, E. E. F. Ballantyne, and D. A. Stone, "Rural EV charging: The effects of charging behaviour and electricity tariffs," *Energy Rep.*, vol. 9, pp. 2321–2334, 2023.

[23] G. Falchetta and M. Noussan, "Electric vehicle charging network in Europe: An accessibility and deployment trends analysis," *Transp. Res. Part D: Transp. Environ.*, vol. 94, pp. 1–18, 2021.

[24] T. Zema, A. Sulich, and S. Grzesiak, "Charging stations and electromobility development: A cross-country comparative analysis," *Energies*, vol. 16, no. 1, pp. 1–20, 2023.

[25] A. Esmaili, M. M. Oshanreh, S. Naderian, D. MacKenzie, and C. Chen, "Assessing the spatial distributions of public electric vehicle charging stations with emphasis on equity considerations in King County, Washington," *Sustain. Cities Soc.*, vol. 107, pp. 1–38, 2024.

[26] H. A. U. Khan, S. Price, C. Avraam, and Y. Dvorkin, "Inequitable access to EV charging infrastructure," *Electr. J.*, vol. 35, no. 3, pp. 1–5, 2022.

[27] C.-W. Hsu and K. Fingerman, "Public electric vehicle charger access disparities across race and income in California," *Transp. Policy*, vol. 100, pp. 59–67, 2021.

[28] J. Vega-Perkins, J. P. Newell, and G. Keoleian, "Mapping electric vehicle impacts: greenhouse gas emissions, fuel costs, and energy justice in the United States," *Environ. Res. Lett.*, vol. 18, no. 1, pp. 1–18, 2023.

[29] US Census Bureau, *Gazetteer Files*, https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html, 2025, retrieved: October, 2025.

[30] US Census Bureau, *Census Datasets*, https://www.census.gov/data/datasets.html, 2025, retrieved: October, 2025.

[31] US Department of Energy, *Alternative Fuels Data Center: Data Downloads*, https://afdc.energy.gov/data_download, 2025, retrieved: October, 2025.

[32] R Core Team, *R: A Language and Environment for Statistical Computing*, Version 4.5.0 (2025-04-11 ucrt), Data analysis software. Available: https://www.R-project.org/, 2025, retrieved: November, 2025.

[33] T. Jonas, O. Okele, and G. A. Macht, "Rural vs. urban: How urbanicity shapes electric vehicle charging behavior in Rhode Island," *World Electr. Veh.*, vol. 16, no. 1, pp. 1–23, 2025.

[34] P. O'Connor, B. Mandel, D. Welch, A. Bolduc, and P. Stith, "Evaluating electric vehicle infrastructure in New Hampshire," New Hampshire Department of Business and Economic Affairs, Tech. Rep., 2019. [Online]. Available: https://www.des.nh.gov/sites/g/files/ehbemt341/files/documents/2020-01/20190524-nh-ev-infrastructure-analysis.pdf, retrieved: October, 2025.

[35] F. Wagner, J. Francfort, and S. White, "Massachusetts plug-in electric vehicle and charging infrastructure case study," Idaho National Laboratory, Tech. Rep. INL/EXT-16-40363, 2016. [Online]. Available: https://www.researchgate.net/publication/315114549_Massachusetts_Plug-in_Electric_Vehicle_and_Charging_Infrastructure_Case_Study, retrieved: October, 2025.

[36] C. Halter, *A record year for electric and plug-in hybrid vehicles in washington*, Department of Ecology, State of Washington, April 2024. [Online]. Available: https://ecology.wa.gov/blog/april-2024/a-record-year-for-electric-vehicles-and-plug-in-hybrids-in-washington, retrieved: October, 2025.

[37] V. McConnell, B. Leard, and F. Kardos, "California's evolving zero emission vehicle program: Pulling new technology into the market," New Hampshire Department of Business and Economic Affairs, Tech. Rep. Working Paper 19–22, 2019. [Online]. Available: https://media.rff.org/documents/RFF_WP_Californias_Evolving_Zero_Emission_Vehicle_Program.pdf, retrieved: October, 2025.

[38] E. M. Hennessy and S. M. Syal, "Assessing justice in California's transition to electric vehicles," *iScience*, vol. 26, no. 7, pp. 1–17, 2023.

[39] C. Ledna, M. Muratori, A. Brooker, E. Wood, and D. Greene, "How to support EV adoption: Tradeoffs between charging infrastructure investments and vehicle subsidies in California," *Energy Policy*, vol. 165, pp. 1–42, 2022.