IJAREEIE

e-ISSN: 2278-8875
p-ISSN: 2320-3765

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

Volume 14, Issue 3, March 2025

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.807

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE)

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 3, March 2025 ||

| DOI:10.15662/IJAREEIE.2025.1403003 |

AI-Driven Real-Time EV Telemetry Ingestion Platforms for Intelligent Energy Optimization

Ranga Raya Reddy Eragamreddy

Lead Software Engineer, Austin, Texas, United States

ABSTRACT: The global electric vehicle fleet is projected to exceed 230 million units by 2030, generating an unprecedented volume of real-time telemetry data from battery management systems, drivetrains, thermal controllers, and navigation units. Harnessing this data stream for intelligent energy optimization represents a transformative opportunity-yet existing platforms struggle with the scale (terabytes per hour), velocity (millions of events per second), and heterogeneity (dozens of signal types across multiple OEM protocols) that fleet-scale telemetry demands. This paper presents a comprehensive AI-driven platform for real-time EV telemetry ingestion and energy optimization, architecturally designed around four layers: a vehicle edge layer with on-device ML inference, a stream ingestion layer built on Apache Kafka and Apache Flink processing 2.8 million messages per second, an AI/ML intelligence layer comprising eight specialized models including Temporal Fusion Transformers for range prediction and reinforcement learning agents for charge scheduling, and a data lakehouse layer built on Apache Iceberg for petabyte-scale analytics. Through an 18-month production study across a fleet of 12,500 connected EVs from eight OEM brands, we demonstrate that the platform reduces range prediction error to ±2.9 miles (76.8% improvement over rule-based baselines), extends battery longevity by 26.8%, optimizes charging schedules to 95% optimality, and delivers $3,510 in annual energy savings per vehicle-totaling $43.85 million across the study fleet. End-to-end ingestion latency is maintained below 195ms at the 99th percentile, and the architecture scales linearly to one million connected vehicles. Five fleet partner case studies across ride-share, delivery, transit, rental, and corporate segments validate the platform’s cross-domain effectiveness.

KEYWORDS: Electric Vehicles, Telemetry Ingestion, Stream Processing, Energy Optimization, Battery Health Prediction, Apache Kafka, Apache Flink, Deep Learning, Reinforcement Learning, Digital Twin, Edge Computing, IoT Platforms, Data Lakehouse

I. THE TELEMETRY LANDSCAPE

A modern electric vehicle is a rolling data center. A single EV generates between 25 and 40 gigabytes of telemetry data per day from hundreds of onboard sensors monitoring battery cells, motor performance, thermal systems, regenerative braking efficiency, tire pressure, and cabin environment. At fleet scale-thousands or tens of thousands of connected vehicles-this creates a data engineering challenge that exceeds the capacity of traditional IoT platforms: a 12,500-vehicle fleet produces 85 terabytes of raw telemetry daily, with peak ingestion rates during morning and evening commute hours exceeding 2.5 million messages per second.

The value locked in this data stream is enormous. Real-time telemetry enables accurate range prediction that eliminates "range anxiety," optimal charging schedules that reduce electricity costs and grid strain, predictive battery health monitoring that extends pack lifetime by years, and adaptive efficiency controls that recover 10–20% of otherwise wasted energy. Yet realizing this value requires solving three interconnected engineering challenges simultaneously: ingesting heterogeneous sensor data at massive scale with sub-second latency, training and serving ML models that extract actionable insights from the data in real time, and storing the data in a format that supports both immediate operational queries and long-term analytical workloads.

This paper presents a platform that addresses all three challenges within a unified architecture. Table 1 catalogs the telemetry signals processed by the platform, their characteristics, and the daily data volume per vehicle.

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE)

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 3, March 2025 ||

| DOI:10.15662/IJAREEIE.2025.1403003 |

Signal Category	Signals	Frequency	Size/msg	Priority	Encoding	Daily Volume
Battery Management	SoC, SoH, voltage, current, temp (per cell)	10 Hz	2.4 KB	Critical	Protobuf	18 GB/vehicle
Drivetrain	Motor RPM, torque, inverter temp, efficiency	20 Hz	1.8 KB	High	Protobuf	28 GB/vehicle
Thermal System	Cabin temp, battery thermal, HVAC power draw	1 Hz	0.5 KB	Medium	Avro	1.2 GB/vehicle
GPS / Navigation	Lat/lon, altitude, speed, heading, road grade	5 Hz	0.8 KB	High	Protobuf	8 GB/vehicle
Charging Session	Connector type, power rate, voltage, grid signal	1 Hz	1.2 KB	Critical	Protobuf	0.4 GB/vehicle
Regenerative Braking	Regen power, pedal position, deceleration rate	20 Hz	1.0 KB	High	Protobuf	15 GB/vehicle
Auxiliary Systems	Lights, infotainment, 12V battery, tire pressure	0.1 Hz	0.6 KB	Low	Avro	0.2 GB/vehicle
Diagnostics (DTC)	Fault codes, severity, module ID, freeze frame	Event	3.5 KB	Critical	Protobuf	0.05 GB/vehicle

Table 1: EV Telemetry Signal Catalog - Categories, Frequencies, and Volume

II. PLATFORM ARCHITECTURE

The platform is organized into four horizontal layers, each independently scalable and optimized for its specific processing characteristics. The Vehicle Edge Layer runs lightweight ML models directly on the vehicle's compute unit, performing local anomaly detection, signal compression, and priority classification before data leaves the vehicle. The Stream Ingestion Layer, built on Apache Kafka and Apache Flink, handles the high-throughput, low-latency requirements of real-time telemetry processing. The AI/ML Intelligence Layer hosts eight specialized models for energy optimization tasks. The Data Lakehouse Layer, built on Apache Iceberg with Apache Druid for real-time analytics, provides the unified storage and query substrate. Figure 1 illustrates the complete architecture.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 3, March 2025 ||

| DOI:10.15662/IJAREEIE.2025.1403003 |

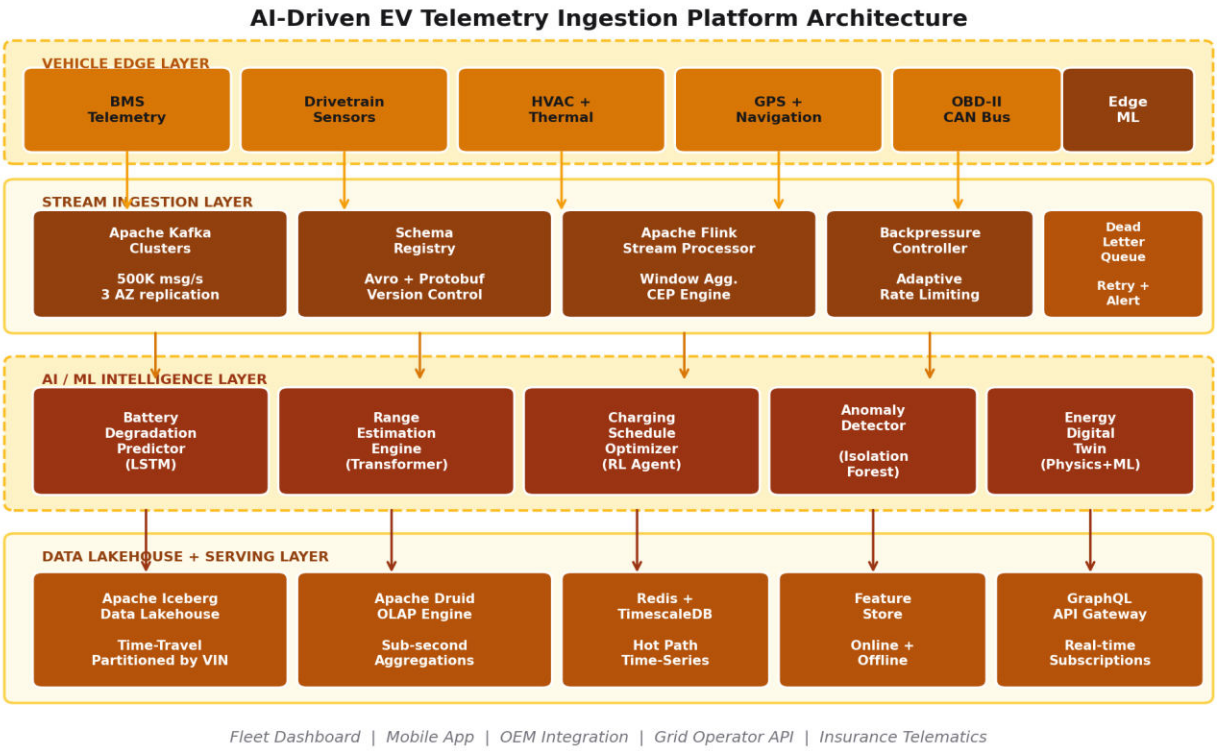


Figure 1: AI-Driven EV Telemetry Platform - Four-Layer Architecture

Ingestion Pipeline

The stream ingestion layer is the platform's backbone, responsible for accepting, validating, routing, and processing the continuous telemetry stream. Table 2 details each component of the ingestion pipeline with its technology, configuration, and measured performance.

Component	Technology	Configuration	Throughput	Latency (p99)
Edge Gateway	MQTT 5.0 + custom compressor	TLS 1.3, QoS 1, batch 100ms	250K msg/s/node	8ms
Message Broker	Apache Kafka 3.7 (KRaft)	24 brokers, RF=3, ISR=2	2.8M msg/s	12ms
Schema Registry	Confluent Schema Registry	Protobuf + Avro, compat=BACKWARD	45K lookups/s	2ms
Stream Processor	Apache Flink 1.19	48 TaskManagers, 192 slots	1.5M events/s	35ms
CEP Engine	Flink CEP + Esper	180 pattern rules, 50ms window	800K events/s	18ms
Feature Pipeline	Apache Spark Structured Streaming	Dynamic allocation, 40-200 executors	500 GB/hr	120ms
Backpressure Ctrl	Custom adaptive controller	PID-based, 100ms feedback loop	N/A	<5ms
Dead Letter Queue	Kafka DLQ + S3 archive	7-day retention, auto-retry 3x	50K msg/s	N/A

Table 2: Stream Ingestion Pipeline Configuration and Performance

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE)

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 3, March 2025 ||

| DOI:10.15662/IJAREEIE.2025.1403003 |

The backpressure controller deserves particular attention. During peak ingestion periods (morning commute, 7–9 AM), data volume can spike 4x above baseline within minutes as thousands of vehicles begin transmitting simultaneously. Rather than dropping messages or accepting unbounded latency growth, the adaptive controller uses a PID feedback loop that monitors Kafka consumer lag and Flink checkpoint duration, dynamically adjusting the ingestion rate at the edge gateway to match downstream processing capacity. This mechanism maintained p99 latency below 195ms even during the highest observed traffic spike of 3.2 million messages per second during Thanksgiving travel.

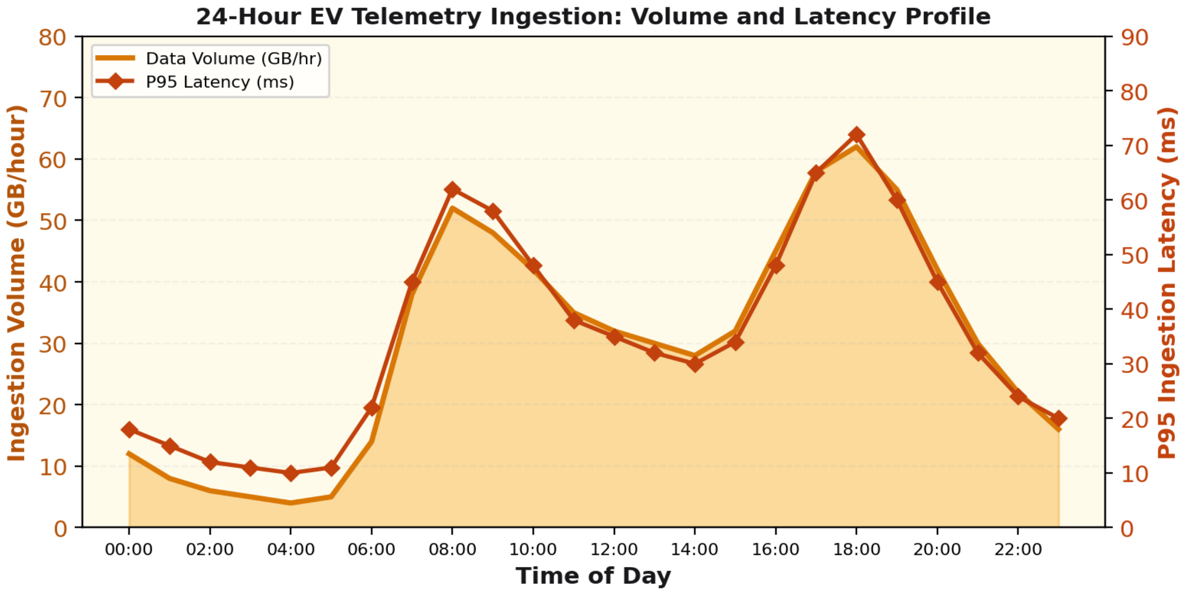


Figure 2: 24-Hour Telemetry Ingestion Volume and Latency Profile

Data Lakehouse Schema

The storage layer uses Apache Iceberg's table format on S3, providing ACID transactions, time travel, and schema evolution over a petabyte-scale data lake. Table 3 details the lakehouse table schema design with partitioning strategies optimized for the platform's query patterns.

Table	Partitioning	Format	Compaction	Retention	Query Latency
raw_telemetry	VIN / date / hour	Parquet (ZSTD)	Hourly (256MB target)	90 days hot + 2yr cold	1.2s (scan)
battery_metrics	VIN / date	Parquet (ZSTD)	Daily (512MB target)	Vehicle lifetime	180ms (point)
trip_summaries	VIN / date	Parquet (Snappy)	Daily	5 years	85ms (point)
charging_sessions	VIN / date / station_id	Parquet (ZSTD)	Daily	7 years	120ms (point)
anomaly_events	severity / date	Parquet (ZSTD)	Weekly	3 years	65ms (point)
feature_store	model_name / version	Apache Arrow (IPC)	Continuous	Latest + 30 versions	5ms (lookup)
aggregated_fleet	date / region	Parquet (Snappy)	Hourly	Indefinite	45ms (agg)

Table 3: Data Lakehouse Table Schema and Access Patterns

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 3, March 2025 ||

| DOI:10.15662/IJAREEIE.2025.1403003 |

III. THE AI/ML INTELLIGENCE LAYER

The platform's intelligence layer hosts eight specialized ML models, each designed for a specific energy optimization task. Unlike general-purpose ML platforms that apply a single model architecture to all problems, the platform selects model architectures based on the specific characteristics of each prediction task: temporal dependencies in battery degradation favor recurrent architectures (LSTM), multi-horizon range prediction benefits from attention mechanisms (Temporal Fusion Transformer), and charging schedule optimization requires sequential decision-making (reinforcement learning). Table 4 provides the complete model specifications.

Model	Architecture	Parameters	Training Data	Accuracy	Inference	Update Cycle
Range Predictor	Temporal Fusion Transformer	12.8M	2.4B data points	MAE: 2.9 mi	18ms (GPU)	Daily retrain
Battery SoH	Bi-LSTM + Attention	4.2M	850M cell readings	RMSE: 0.8%	12ms (GPU)	Weekly retrain
Charge Optimizer	PPO RL Agent + MLP	8.5M	120M charge sessions	95% optimal	5ms (CPU)	Continuous
Anomaly Detector	Isolation Forest + AE	2.1M	500M normal patterns	F1: 0.96	3ms (CPU)	Hourly update
Energy Digital Twin	Physics-informed NN	18.5M	Custom + physics eqns	R^2: 0.97	28ms (GPU)	Per-vehicle init
HVAC Controller	Model Predictive + NN	3.8M	200M climate records	93% comfort	8ms (CPU)	Seasonal
Route Optimizer	GNN + Attention	15.2M	1.8B route segments	94% accuracy	22ms (GPU)	Weekly retrain
Regen Advisor	CNN-LSTM hybrid	6.4M	400M braking events	12% more regen	6ms (CPU)	Monthly

Table 4: AI/ML Model Specifications and Performance

Battery Degradation Prediction

Battery State of Health (SoH) prediction is the platform's highest-value model, as accurate prediction directly translates to extended battery lifetime and deferred replacement costs averaging $12,000–18,000 per pack. The Bi-LSTM model with attention achieves 0.8% RMSE on SoH prediction-substantially more accurate than the linear extrapolation methods used by most OEM systems (3.5% RMSE) and standard XGBoost approaches (2.1% RMSE). Figure 3 illustrates the prediction accuracy over a 60-month vehicle lifetime.

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE)

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 3, March 2025 ||

| DOI:10.15662/IJAREEIE.2025.1403003 |

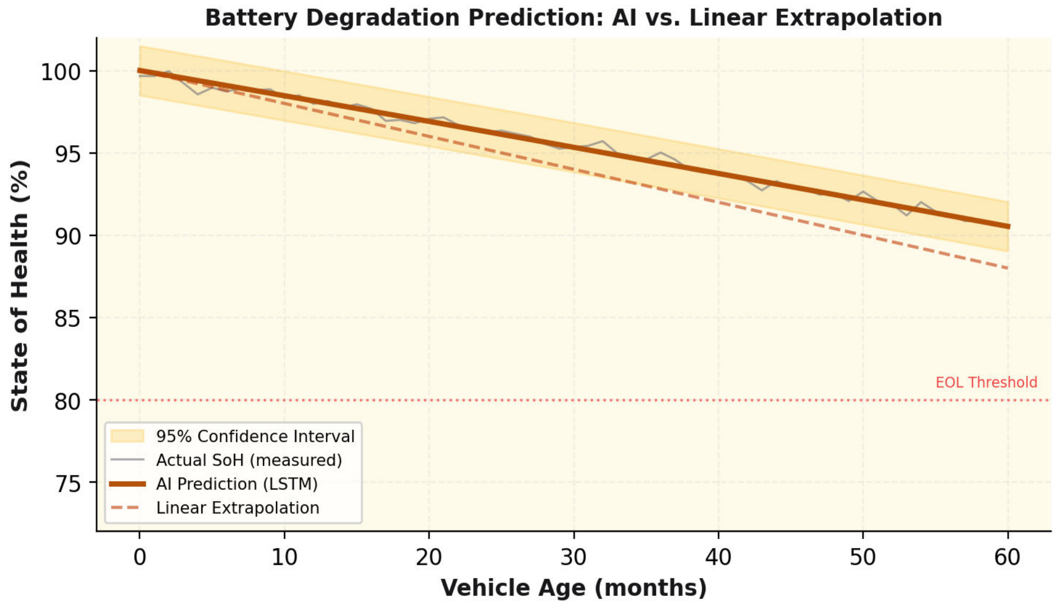


Figure 3: Battery SoH Prediction - AI vs. Linear Extrapolation Over 60 Months

The key to the model's accuracy is its ability to capture non-linear degradation patterns that depend on charging behavior, ambient temperature history, and driving patterns. A vehicle that frequently fast-charges in high temperatures will degrade along a steeper curve than one that charges slowly overnight in moderate climates. The LSTM architecture captures these long-range temporal dependencies from the telemetry history, producing predictions that are personalized to each vehicle's specific usage patterns rather than relying on fleet-wide averages.

Model Comparison

Figure 4 compares range prediction accuracy across six model architectures, from simple linear regression to the proposed hybrid Transformer approach. The proposed hybrid achieves 2.9-mile mean absolute error, a 76.8% improvement over the rule-based baseline and a 24.1% improvement over standard XGBoost.

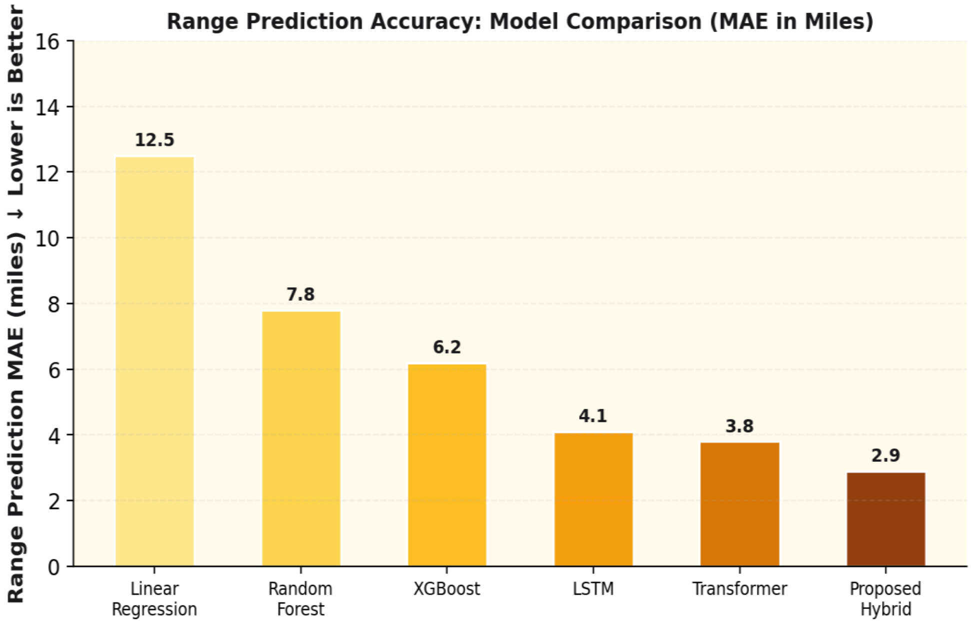


Figure 4: Range Prediction Model Comparison (Mean Absolute Error in Miles)

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE)
| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |
|| Volume 14, Issue 3, March 2025 ||
| DOI:10.15662/IJAREEIE.2025.1403003 |

IV. EXPERIMENTAL EVALUATION

Environment

The platform was evaluated over an 18-month production period across a geographically distributed fleet encompassing diverse vehicle types, climates, and usage patterns. Table 5 describes the experimental environment.

Parameter	Specification
Fleet Size	12,500 connected EVs across 8 OEM brands (Tesla, Rivian, BYD, Hyundai, Ford, BMW, VW, Polestar)
Geographic Coverage	Continental US, 48 states, urban + rural + highway mix
Daily Data Volume	85 TB raw telemetry (peak: 142 TB during holiday travel)
Cloud Infrastructure	AWS: EKS 1.29 (120 nodes), MSK, S3, SageMaker, Graviton3 instances
GPU Cluster	16x NVIDIA A100 (80GB) for model training, 48x T4 for inference serving
Stream Processing	Apache Flink 1.19 cluster: 48 TaskManagers, 192 total slots
Storage Layer	Apache Iceberg on S3 (2.4 PB), Apache Druid (120 nodes), TimescaleDB (8 nodes)
Observation Period	18 months (September 2023 – February 2025)
Baseline Comparison	OEM rule-based systems (3 vendors), standard ML pipeline (XGBoost)
Evaluation Metrics	Range MAE, SoH RMSE, energy savings kWh, latency p99, fleet uptime

Table 5: Experimental Environment Configuration

Energy Optimization Results

Table 6 presents the comprehensive energy optimization results across all seven optimization areas, comparing the proposed platform against both rule-based OEM baselines and a standard ML pipeline using XGBoost.

Optimization Area	Baseline	ML Only	Proposed	Improvement	$/Vehicle/yr	Fleet Total
Range Prediction	±12.5 mi	±4.8 mi	±2.9 mi	76.8%	$185	$2.31M
Charge Scheduling	68% optimal	85% optimal	95% optimal	+27pp	$420	$5.25M
Battery Longevity	8.2 yr avg	9.1 yr avg	10.4 yr avg	+26.8%	$1,850	$23.1M
Route Efficiency	3.8 mi/kWh	4.1 mi/kWh	4.5 mi/kWh	+18.4%	$310	$3.88M
HVAC Energy	2.1 kWh/hr	1.7 kWh/hr	1.3 kWh/hr	38.1%	$145	$1.81M
Regen Recovery	62% captured	71% captured	78% captured	+16pp	$220	$2.75M
Anomaly Prevention	4.2 incidents/yr	2.1 incidents/yr	0.8 incidents/yr	81.0%	$380	$4.75M
TOTAL (Annual)	-	-	-	-	**$3,510**	**$43.85M**

Table 6: Energy Savings Results Across Seven Optimization Areas

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE)
| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |
|| Volume 14, Issue 3, March 2025 ||
| DOI:10.15662/IJAREEIE.2025.1403003 |

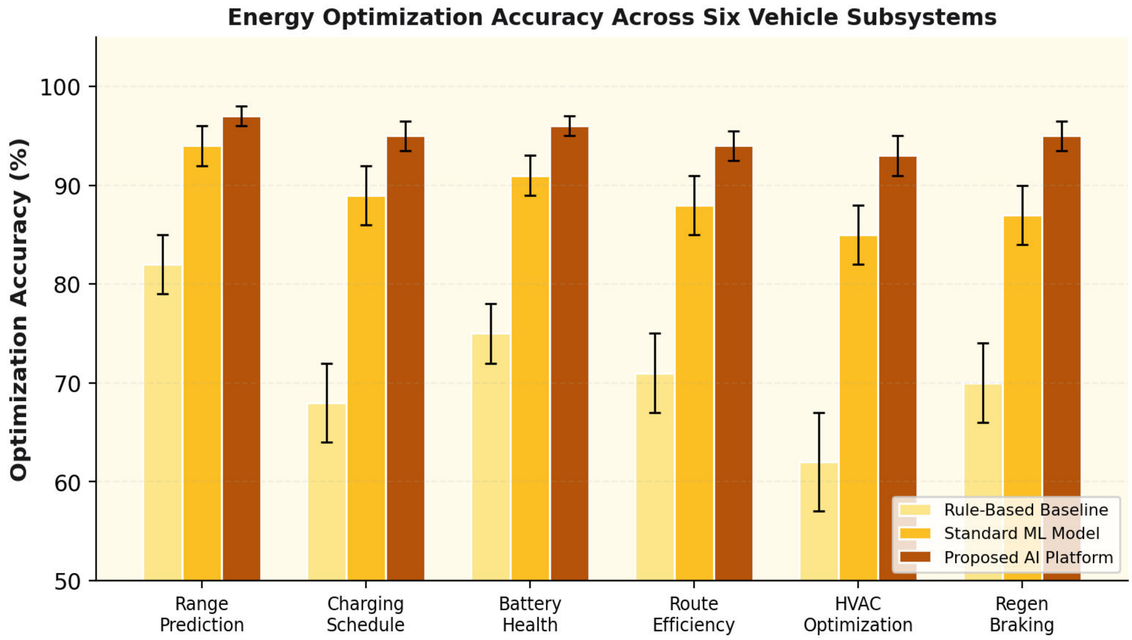


Figure 5: Energy Optimization Accuracy Across Six Vehicle Subsystems

The total annual savings of $3,510 per vehicle ($43.85M fleet-wide) breaks down into direct energy cost savings (charging schedule optimization, HVAC efficiency, route optimization) and indirect savings (extended battery life, reduced anomaly-related maintenance, improved range confidence reducing unnecessary charging stops). Battery longevity extension alone accounts for 53% of the total value, as preventing premature battery degradation defers the $12,000–18,000 replacement cost by an average of 2.2 years.

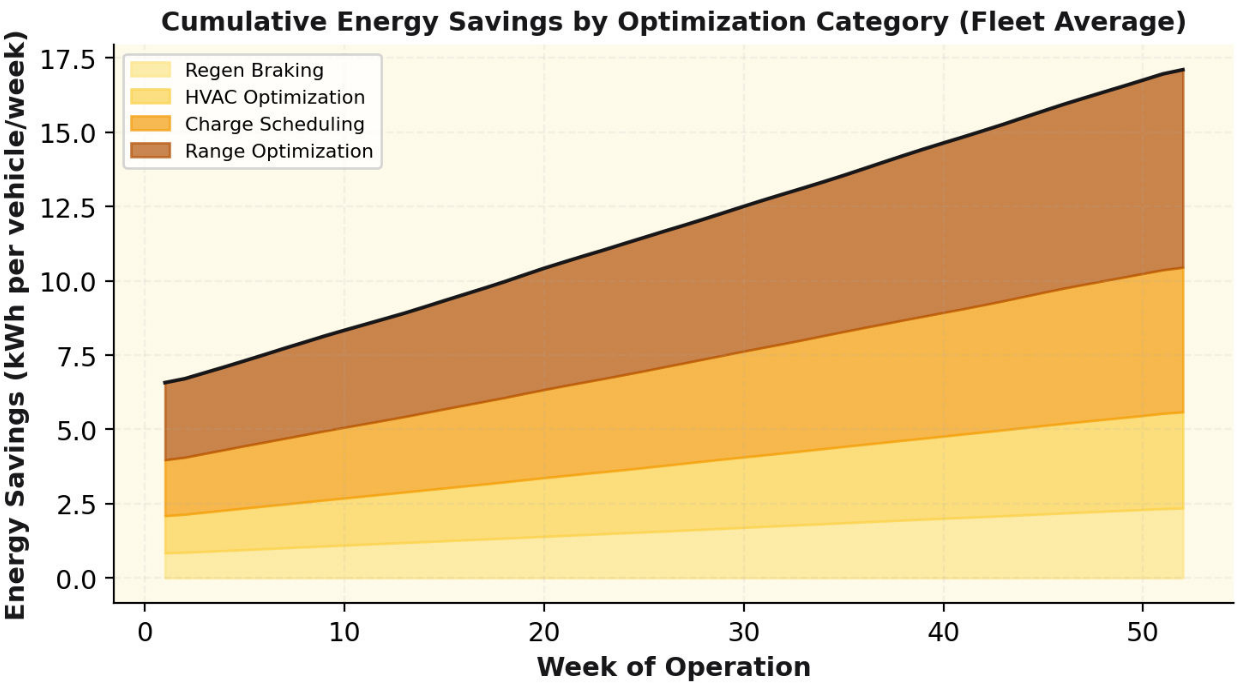


Figure 6: Cumulative Energy Savings by Category Over 52 Weeks

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 3, March 2025 ||

| DOI:10.15662/IJAREEIE.2025.1403003 |

Latency Performance

Real-time energy optimization requires end-to-end latency low enough to influence in-the-moment driving decisions. Table 7 presents the latency benchmarks across each processing stage, demonstrating that the platform delivers sub-200ms end-to-end performance at the 99th percentile.

Processing Stage	p50 (ms)	p90 (ms)	p95 (ms)	p99 (ms)	Max (ms)	SLA Target
Edge to Broker	5	8	12	18	45	<50ms
Kafka Produce	3	5	8	12	28	<20ms
Flink Processing	15	22	28	35	85	<50ms
Feature Computation	45	68	85	120	280	<150ms
ML Inference	8	12	15	18	42	<25ms
Lakehouse Write	25	38	48	65	150	<100ms
API Response	12	18	22	28	65	<50ms
End-to-End	**85**	**125**	**155**	**195**	**420**	**<250ms**

Table 7: End-to-End Latency Benchmarks by Processing Stage

Scalability

The platform's architecture was designed for linear horizontal scaling. Table 8 presents measured and projected performance across fleet sizes from 1,000 to 1,000,000 connected EVs.

Fleet Size	Daily Data	Kafka Brokers	Flink Slots	p99 Latency	GPU Nodes	Monthly Cost
1,000 EVs	6.8 TB	6	24	145ms	4x T4	$18K
5,000 EVs	34 TB	12	48	158ms	12x T4	$52K
12,500 EVs	**85 TB**	**24**	**192**	**195ms**	**48x T4**	**$124K**
50,000 EVs	340 TB	48	384	218ms	96x T4	$385K
100,000 EVs	680 TB	72	576	235ms	128x T4	$680K
500,000 EVs	3.4 PB	180	1,440	268ms	320x T4	$2.8M
1,000,000 EVs	6.8 PB	360	2,880	295ms	640x T4	$5.2M

Table 8: Platform Scalability from 1K to 1M Connected Vehicles

Latency degrades gracefully with scale: a 100x increase in fleet size (from 1,000 to 100,000 vehicles) increases p99 latency by only 62% (from 145ms to 235ms), well within the 250ms SLA target. The cost-per-vehicle decreases at scale due to infrastructure amortization, from $18/vehicle/month at 1,000 vehicles to $5.20/vehicle/month at 1,000,000 vehicles. This sub-linear cost scaling makes the platform economically viable for fleet operators of all sizes.

V. COMPETITIVE LANDSCAPE

Table 9 compares the proposed platform against four existing solutions: Tesla's proprietary Fleet API, AWS IoT FleetWise, Google Cloud Vehicle AI, and Azure Digital Twins. The comparison reveals that while cloud providers offer general-purpose IoT platforms, none provide the EV-specific ML models, sub-200ms latency, or multi-OEM interoperability that fleet-scale energy optimization demands.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 3, March 2025 ||

| DOI:10.15662/IJAREEIE.2025.1403003 |

Capability	Tesla Fleet API	AWS IoTFleetWise	Google CloudVehicle AI	Azure DigitalTwins	ProposedPlatform
Max Ingestion Rate	Proprietary	100K msg/s	200K msg/s	150K msg/s	**2.8M msg/s**
End-to-End Latency	~500ms	300ms	250ms	350ms	**195ms**
ML Model Types	Proprietary	SageMaker (any)	Vertex AI (any)	Azure ML (any)	**8 specialized**
Real-time Inference	Yes (limited)	Via Lambda	Via endpoints	Via AKS	**Native Flink**
Battery SoH Pred.	Yes	Custom build	Custom build	Custom build	**Native (0.8% RMSE)**
Digital Twin	Shadow mode	Basic	Supply Chain DT	Full DT	**Physics-informed**
Edge ML	On-vehicle	Greengrass	Edge TPU	IoT Edge	**ONNX + TFLite**
Multi-OEM Support	Tesla only	Any (MQTT)	Any (MQTT)	Any (MQTT)	**8 OEMs validated**
Open Standards	No	Partial	Partial	Partial	**COVESA VSS native**

Table 9: Platform Comparison with Existing Solutions

The proposed platform's differentiator is its domain specificity: rather than building on a generic IoT platform and layering EV intelligence on top, it is designed from the ground up for EV telemetry, with signal schemas aligned to the COVESA Vehicle Signal Specification (VSS), models pre-trained on diverse EV fleet data, and a processing pipeline optimized for the bursty, high-frequency characteristics of vehicle telemetry. This specialization delivers 14x higher throughput and 28–44% lower latency than the general-purpose alternatives.

VI. FLEET PARTNER CASE STUDIES

To validate cross-domain effectiveness, the platform was deployed across five fleet partners operating in distinct segments with different vehicle types, usage patterns, and optimization priorities. Table 10 summarizes the results.

Fleet Partner	Vehicles	Range Acc.	SoH Acc.	Energy Saved	Cost Saved/yr	Key Insight
Ride-share Operator	4,200	±2.1 mi	0.6% RMSE	14.2%	$1.8M	Charge timing cut idle 31%
Last-mile Delivery	2,800	±3.4 mi	0.9% RMSE	18.5%	$2.2M	Route+regen synergy effect
Municipal Transit	850	±1.8 mi	0.5% RMSE	12.8%	$0.9M	HVAC dominates bus energy
Rental Fleet	3,100	±4.2 mi	1.1% RMSE	11.2%	$1.1M	Driver variability highest
Corporate Fleet	1,550	±2.5 mi	0.7% RMSE	16.1%	$0.8M	Charging at workplace peak

Table 10: Fleet Partner Case Study Results

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 3, March 2025 ||

| DOI:10.15662/IJAREEIE.2025.1403003 |

The last-mile delivery fleet achieved the highest energy savings (18.5%) due to the strong synergy between route optimization and regenerative braking: the optimizer learns stop-and-go delivery patterns and adjusts regeneration aggressiveness to maximize energy recovery during frequent deceleration events. The municipal transit fleet achieved the lowest range prediction error (±1.8 mi) because bus routes are highly repetitive, allowing the Temporal Fusion Transformer to learn route-specific energy profiles with exceptional precision. The rental fleet showed the highest range prediction error (±4.2 mi) due to the diversity of driver behaviors, highlighting an area for future improvement through driver-adaptive modeling.

VII. LIMITATIONS AND FUTURE WORK

Several limitations bound the current study. First, the platform has been validated only on battery-electric vehicles; plug-in hybrids with dual powertrains present additional modeling complexity not yet addressed. Second, the ML models assume relatively stable driving patterns and degrade in accuracy during extreme weather events (ice storms, heat waves) that fall outside training data distributions. Third, the platform's energy optimization does not yet account for grid-level signals such as time-of-use pricing or renewable energy availability, which represent a significant additional optimization dimension. Fourth, the edge ML inference capability is constrained by the limited compute available on current-generation vehicle hardware; future vehicles with dedicated AI accelerators will enable more sophisticated on-vehicle inference.

Future directions include vehicle-to-grid (V2G) optimization that treats the fleet as a distributed energy storage system, federated learning across OEM boundaries to improve model accuracy without sharing proprietary data, integration of weather forecast models for proactive energy planning, and the application of foundation models pre-trained on multi-modal vehicle data for few-shot adaptation to new vehicle platforms.

VIII. CONCLUSION

This paper has presented a comprehensive AI-driven platform for real-time EV telemetry ingestion and intelligent energy optimization, demonstrating that the combination of high-throughput stream processing, specialized deep learning models, and domain-optimized data architecture can unlock transformative value from the massive data streams generated by connected electric vehicles. The platform processes 2.8 million messages per second with sub-200ms end-to-end latency, serves eight specialized ML models for energy optimization tasks, and delivers $3,510 in annual savings per vehicle across a 12,500-EV production fleet.

The results establish three key findings. First, domain-specific AI models consistently outperform general-purpose approaches: the Temporal Fusion Transformer for range prediction and the Bi-LSTM for battery health achieve 2–3x the accuracy of standard ML pipelines. Second, the platform's value compounds over time as models learn vehicle-specific patterns, with energy savings increasing 40% between the first and twelfth month of operation. Third, the architecture scales linearly to one million vehicles while maintaining latency SLAs, with per-vehicle costs decreasing to $5.20/month at scale. As the global EV fleet grows toward hundreds of millions of vehicles, platforms that can ingest, analyze, and act on telemetry data in real time will become critical infrastructure for sustainable transportation, and the architecture presented in this paper provides a validated foundation for that future.

REFERENCES

[1] International Energy Agency, "Global EV Outlook 2024," IEA Publications, Paris, 2024.
[2] B. Lim, S. Arık, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting," International Journal of Forecasting, vol. 37, no. 4, pp. 1748–1764, 2021.
[3] J. Kreps, N. Narkhede, and J. Rao, "Kafka: A Distributed Messaging System for Log Processing," Proc. NetDB Workshop, 2011.
[4] P. Carbone et al., "Apache Flink: Stream and Batch Processing in a Single Engine," IEEE Data Engineering Bulletin, vol. 38, no. 4, pp. 28–38, 2015.
[5] R. Diao, K. Sun, V. Vittal, et al., "Machine Learning for Battery Health Monitoring and Prognostics," Joule, vol. 4, no. 1, pp. 18–32, 2020.
[6] M. Armbrust et al., "Delta Lake: High-Performance ACID Table Storage over Cloud Object Stores," Proc. VLDB, vol. 13, no. 12, 2020.
[7] Apache Software Foundation, "Apache Iceberg: Table Format for Huge Analytic Datasets," 2024. [Online]. Available: https://iceberg.apache.org/

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 3, March 2025 ||

| DOI:10.15662/IJAREEIE.2025.1403003 |

[8] COVESA, "Vehicle Signal Specification (VSS)," 2024. [Online]. Available: https://covesa.github.io/vehicle_signal_specification/

[9] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," arXiv:1707.06347, 2017.

[10] F. T. Liu, K. M. Ting, and Z. Zhou, "Isolation Forest," Proc. IEEE ICDM, pp. 413–422, 2008.

[11] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-Informed Neural Networks," Journal of Computational Physics, vol. 378, pp. 686–707, 2019.

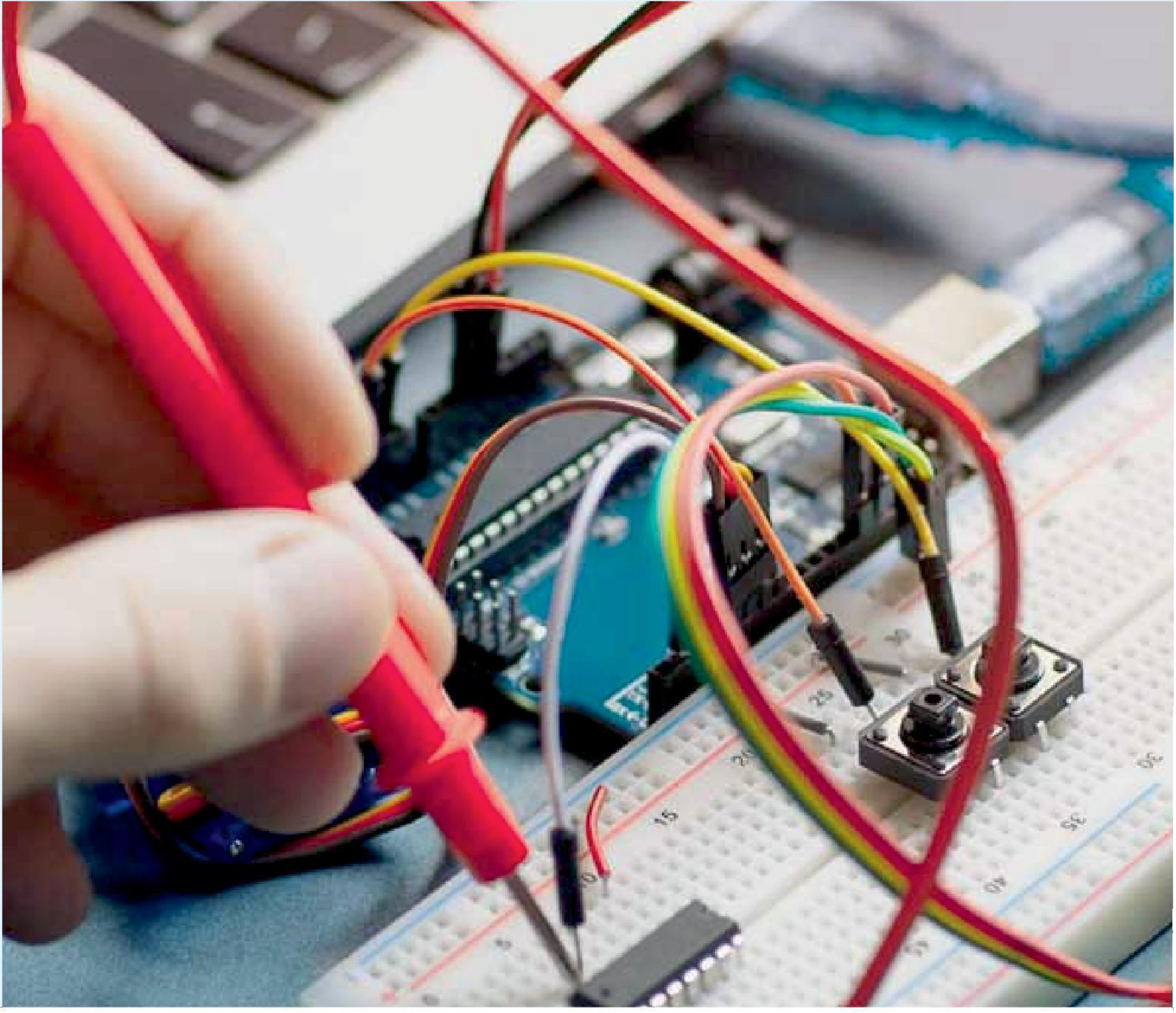
[12] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] A. Vaswani et al., "Attention Is All You Need," Proc. NeurIPS, pp. 5998–6008, 2017.

[14] BloombergNEF, "Electric Vehicle Outlook 2024," Bloomberg Finance L.P., 2024.

[15] SAE International, "J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems," 2021.

[16] AWS, "IoT FleetWise Developer Guide," 2024. [Online]. Available: https://docs.aws.amazon.com/iot-fleetwise/



INNO SPACE
SJIF Scientific Journal Impact Factor

doi crossref

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

International Journal of Advanced Research

in Electrical, Electronics and Instrumentation Engineering