

OBD-II Big Data-Driven ML- and AI-Based Virtual Sensing for Fuel Economy, Component Health, and Carbon Intelligence

Arumugam Palanichamy^{1*}, Karthikeyan Subramanian²,
Rajavel Rangasamy³, Suresh Alex Selvaraj⁴

Abstract

The rapid growth of connected vehicles has led to the large-scale availability of high-frequency on-board diagnostics II (OBD-II) data; however, much of this data remains underutilized, as existing studies and commercial systems typically address fuel economy, maintenance, or emissions in isolation or rely on additional physical sensors. Such fragmented and sensor-dependent approaches limit scalability and increase system cost, particularly in high-volume and resource-constrained vehicle markets. To address this gap, this study proposes an OBD-II big data-driven machine learning and artificial intelligence based virtual sensing approach that enables simultaneous estimation of fuel economy, component health, and carbon intelligence using only standard on-board signals. Continuous real-world OBD-II data collected during naturalistic vehicle operation are analyzed using physics-guided feature extraction combined with unsupervised learning and interpretable artificial intelligence (AI) techniques to derive sensor-less virtual metrics, including fuel economy (kmpl), gross vehicle weight or payload trends, clutch, and brake wear indicators, battery state-of-health, remaining driving range, and carbon dioxide emissions. By exploiting intrinsic relationships between engine torque demand, vehicle dynamics, operating conditions, and fuel consumption behavior, the proposed approach demonstrates that multiple energy, health, and sustainability indicators can be inferred reliably without additional hardware. Overall, this work supports fuel-efficient operation, proactive maintenance, and accurate carbon monitoring, contributing to reduced fuel consumption, lower greenhouse gas emissions, and improved total cost of ownership, while offering a scalable pathway toward software-defined vehicle intelligence using existing vehicle data infrastructure.

*Author for Correspondence

Arumugam Palanichamy
E-mail: arumugampd@gmail.com

¹Research Scholar, Department of Marine Engineering, AMET University, Chennai, Tamil Nadu, India

²Assistant General Manager, Department of Fuel Cell Development, Ashok Leyland Technical Centre, Chennai, Tamil Nadu, India

³Principal, Head of Engineering Department, Sri Balaji Chockalingam Engineering College, Irumbedu, Arni, Tamil Nadu, India

⁴Professor, Department of Marine Engineering, AMET University, Chennai, Tamil Nadu, India

Received Date: January 19, 2026

Accepted Date: January 25, 2026

Published Date: February 11, 2026

Citation: Arumugam Palanichamy, Karthikeyan Subramanian, Rajavel Rangasamy, Suresh Alex Selvaraj. OBD-II Big Data-Driven ML- and AI-Based Virtual Sensing for Fuel Economy, Component Health, and Carbon Intelligence. Journal of Automobile Engineering and Applications. 2026; 13(1): 41–52p.

Keywords: Carbon emission intelligence, fuel economy analysis, machine learning and artificial intelligence, OBD-II big data, vehicle component health monitoring, virtual sensing

INTRODUCTION

The rapid growth of connected vehicles has resulted in the large-scale availability of high-frequency onboard diagnostics II (OBD-II) data generated during real-world vehicle operations. In this study, such data were used to capture detailed information on vehicle speed, engine behavior, load demand, and fuel usage, highlighting the potential for data-driven vehicle intelligence beyond conventional diagnostics [1–5]. However, despite this growing availability, most existing studies and

commercial implementations continue to utilize OBD-II data in a fragmented manner, treating fuel economy, maintenance, and emissions as isolated objectives rather than interconnected aspects of real-world vehicle operations [6–9].

Traditional vehicle monitoring and maintenance approaches for current production vehicles rely heavily on physical sensors, scheduled inspections, and rule-based thresholds, which increase the system cost, complexity, and maintenance overhead [10–15]. In this context, sensor-dependent architectures limit scalability and widespread adoption, particularly in high-volume and resource-constrained vehicle markets. Motivated by these limitations, the growing emphasis on predictive maintenance, fuel efficiency improvement, and real-world carbon monitoring has highlighted the need for integrated and cost-effective approaches that can directly extract multiple forms of vehicle intelligence from existing onboard data sources [16–25].

Machine learning (ML) and artificial intelligence (AI) techniques have proven effective in analyzing large-scale, nonlinear, and heterogeneous vehicle datasets, thereby enabling the improved prediction and interpretation of vehicle behavior under real-world conditions [1–5, 26]. Prior research has successfully applied ML-based methods to fuel consumption modeling [7, 20, 21], driving behavior analysis [27–31], component health monitoring [15, 18], and battery state-of-health estimation [32, 33]. However, these studies typically focus on individual subsystems or performance metrics, resulting in isolated solutions that do not fully exploit the shared information content of OBD-II data.

Virtual sensing, also known as soft sensing, offers a promising alternative by enabling the estimation of unmeasured or difficult-to-measure variables using data-driven models and physics-based relationships instead of additional physical sensors [11–14]. Although virtual sensing has been widely adopted in industrial process monitoring [11–13], its systematic application in automotive systems remains limited, with most existing studies focusing on single variables rather than the integrated sensing of energy, health, and emissions [14, 29].

In parallel, an accurate assessment of real-world fuel economy and carbon emissions has become increasingly important for regulatory compliance, sustainability assessments, and total cost of ownership optimization [22–24]. Conventional laboratory-based or average-based estimation methods fail to capture the dynamic and context-dependent nature of real-world vehicle operations, highlighting the need for data-driven carbon intelligence derived directly from in-use vehicle data [9, 28, 34, 35].

To address these limitations, this study proposes an OBD-II big data-driven ML- and AI-based virtual sensing approach that enables the simultaneous estimation of fuel economy, component health indicators, and carbon intelligence using only standard onboard signals. By combining physics-guided feature extraction with unsupervised learning and interpretable AI techniques [1, 3, 5], multiple sensor-less virtual metrics, including fuel economy (kmpl), gross vehicle weight or payload trends, clutch and brake wear indicators, battery state-of-health, remaining driving range, and carbon dioxide emissions, can be derived from a unified data source.

The key contributions of this study are threefold: (i) development of a scalable and sensor-independent virtual-sensing methodology using real-world OBD-II big data; (ii) integration of energy efficiency, component health, and carbon emission assessment within a single ML–AI-driven sensing paradigm; and (iii) demonstration of the relevance of virtual sensing for software-defined vehicle intelligence, predictive maintenance, and sustainability-oriented decision support. The proposed approach provides a practical pathway for equipment manufacturers and fleet operators to improve operational efficiency, reduce fuel consumption and emissions, and enhance vehicle intelligence using the existing vehicle data infrastructure.

PROBLEM STATEMENT OVERVIEW

The increasing deployment of connected vehicles has enabled the continuous acquisition of rich operational data using OBD-II systems under real-world driving conditions [6, 8, 9]. However, most

existing research efforts and practical implementations utilize these data in a limited and fragmented manner, typically focusing on individual objectives, such as fuel economy estimation [7, 20, 21], component diagnostics [15, 18], or emission assessment [22, 23]. Such single-metric or subsystem-specific analyses do not fully exploit the inherent coupling between vehicle dynamics, powertrain loading, component degradation, and energy consumption embedded within high-frequency real-world OBD-II data streams [19, 24].

Furthermore, current vehicle monitoring and maintenance strategies continue to rely on dedicated physical sensors, periodic inspections, or predefined rule-based thresholds to assess component condition and operational efficiency [10, 15]. While ML and AI techniques have been successfully applied to selected automotive use cases, including predictive maintenance [16, 17], driving behavior analysis [30, 31], and battery state-of-health estimation [32, 33], these approaches have largely been developed in isolation and target single output variables. Consequently, the potential to infer multiple, interrelated energy, health, and sustainability indicators from a common OBD-II data source using a unified, sensor-independent virtual sensing approach remains underexplored [11–14, 29]. Addressing this gap is essential for enabling scalable software-defined vehicle intelligence that supports real-world fuel efficiency improvement, proactive maintenance, and data-driven carbon monitoring.

Methodology and the Integrated Literature Overview

This study adopted a data-driven methodology that integrated big data analytics, ML, and AI to develop a virtual sensing approach using real-world OBD-II data. The methodology is designed to transform high-frequency vehicle operational data into multiple energy-, health-, and sustainability-related performance indicators without relying on additional physical sensors. This approach aligns with recent trends in automotive analytics that emphasize scalable, software-defined intelligence built on existing on-board data infrastructure [6, 9, 28].

Real-world vehicle data acquired through OBD-II systems exhibit the key characteristics of big data, including high volume, velocity, variety, and variability, requiring systematic preprocessing, feature extraction, and model design [8, 9, 24]. Prior studies have demonstrated the effectiveness of OBD-II data in analyzing fuel consumption behavior [7, 20, 21], driving patterns [30, 31], and emissions under real-world conditions [22, 23]. However, these efforts have largely focused on individual performance metrics, motivating the need for an integrated methodology capable of extracting multiple forms of vehicle intelligence from common data sources.

To address this need, the proposed methodology is structured around the concept of virtual sensing, also referred to as soft sensing, in which unmeasured or difficult-to-measure variables are inferred using data-driven models and physics-guided relationships, rather than additional hardware sensors [11–14]. Virtual sensing approaches have been widely applied in industrial monitoring applications owing to their cost-effectiveness and robustness [11, 12]; however, their systematic application to automotive systems for the simultaneous estimation of energy performance, component health, and carbon emissions remains limited [14, 29].

ML and AI techniques form the analytical core of the proposed methodology, enabling the extraction of meaningful patterns from high-dimensional and nonlinear vehicle data [1–5, 34]. In particular, unsupervised learning methods are well-suited for real-world vehicle datasets, in which labeled fault or performance data are scarce and operating conditions evolve continuously [16, 17]. By combining data-driven learning with physics-based feature construction, the methodology ensures the interpretability and physical consistency of the derived virtual metrics, which are essential for practical automotive deployment [3, 5].

The overall methodology is organized into four logical stages to ensure clarity and reproducibility. First, OBD-II data acquisition and big data characterization are performed to establish data quality and suitability for analysis. Second, the virtual sensing concept and system architecture are defined,

outlining how multiple sensor-less performance indicators can be derived from standard on-board signals. Third, ML–AI model development and feature engineering strategies are described, highlighting the role of physics-guided and unsupervised learning techniques. Finally, the development of individual virtual metrics, including fuel economy, component health indicators, and carbon intelligence, is presented in subsequent sections (Figure 1).

OBD-II Data Acquisition and Big Data Characterization Process

Real-world vehicle operational data were collected using the onboard diagnostics II (OBD-II) interface under normal driving conditions. The OBD-II system provides continuous access to key powertrain, vehicle dynamics, fuel, and electrical parameters, such as vehicle speed, engine speed, engine load, fuel-related signals, and battery voltage, thereby enabling the direct observation of in-use vehicle behavior without reliance on predefined test cycles [6, 8]. Data were recorded over extended driving durations to capture diverse operating conditions and driving patterns of the vehicle.

The acquired OBD-II dataset exhibits essential big data characteristics, including high volume and velocity owing to continuous time-series logging, variety arising from heterogeneous signal types, and variability caused by frequent transitions in traffic, road, and driver behavior [9, 24, 30]. To ensure data veracity, systematic preprocessing was applied, including the removal of communication dropouts, filtering of physically inconsistent values, and consistency checks based on the expected vehicle operating limits [8, 10]. Time synchronization and resampling were performed to maintain uniform temporal resolution across all parameters (Figure 2).

The resulting validated dataset provides a reliable representation of real-world vehicle operations and forms the foundation for subsequent physics-guided feature extraction and ML–AI-based virtual sensing applications. The transformation of processed OBD-II data into meaningful features and their integration within the virtual sensing architecture are described in Section 2.

Virtual Sensing Concept and the System Architecture

Virtual sensing is a data-driven approach that estimates variables that are not directly measured by physical sensors by exploiting existing signals and the underlying physical relationships. In automotive applications, virtual sensing enables the estimation of performance, health, and sustainability-related metrics using standard on-board data, thereby reducing the dependence on additional hardware sensors and improving scalability across vehicle platforms [11–14]. The continuous availability of OBD data during real-world vehicle operations renders this approach particularly suitable for developing software-defined vehicle intelligence.

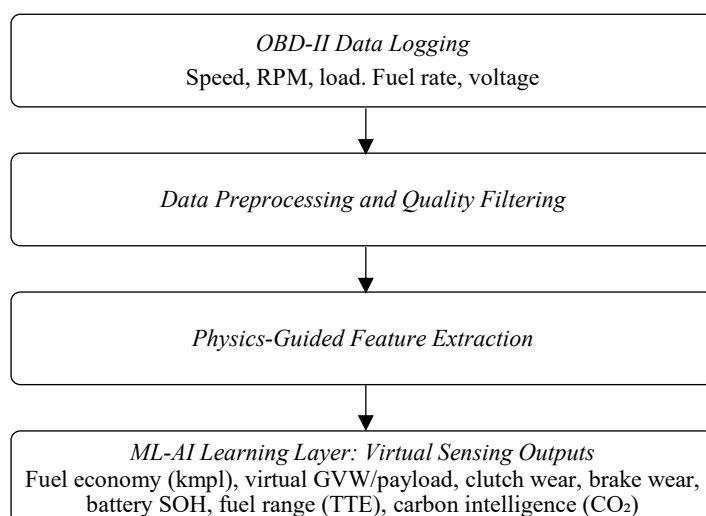


Figure 1. Methodology overview.

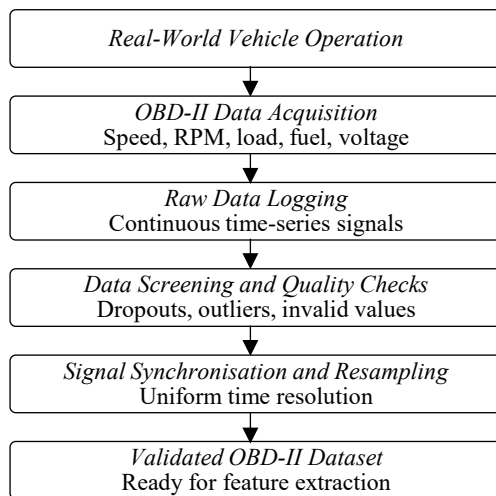


Figure 2. OBD-II Data processing.

The proposed system architecture follows a layered virtual sensing concept in which validated OBD-II data serve as the primary input. In the first stage, physics-guided feature extraction is applied to raw signals such as vehicle speed, engine speed, engine load, fuel-related parameters, and battery voltage. These features are designed to represent meaningful physical quantities related to vehicle dynamics, torque demand, energy usage, and electrical behavior, thereby ensuring interpretability and consistency with the fundamental vehicle operating principles [3, 5, 14].

ML and AI techniques are then used to analyze the extracted features and estimate the virtual sensing outputs. Unsupervised learning methods are particularly suitable for real-world vehicle data, as labeled fault or degradation data are rarely available, and the operating conditions continuously vary during normal driving [16, 17]. By learning the typical operating behavior directly from the data, the ML–AI layer supports the reliable estimation of multiple virtual metrics without relying on predefined rules or fixed thresholds [1, 34].

The virtual sensing outputs include fuel economy, component health indicators, and carbon-related metrics, all of which are derived from the shared set of features. The use of a common feature space allows these outputs to be estimated together, reflecting their natural interdependence during real-world vehicle operations (Figure 3). This integrated approach avoids treating energy efficiency, maintenance, and emissions as separate problems and provides a more practical and scalable solution for vehicle intelligence [7, 15, 22].

ML–AI Model Development and the Feature Engineering (Final Version)

This section describes how validated OBD-II data are transformed into a learning-ready representation and how ML–AI models are developed to support the virtual sensing of fuel economy, component health, and carbon-related metrics. The focus is on learning the normal vehicle operating behavior directly from real-world data to enable reliable sensorless inference.

The engineered features were used as inputs to the ML–AI learning layer. Given the limited availability of labeled fault or degradation data in real-world vehicle operations, unsupervised learning techniques have been employed to learn the characteristic patterns of normal behavior from the data [1, 16, 17]. Rather than predicting a single output variable, the models establish a reference representation of typical vehicle operations across several dimensions. Deviations from this learned behavior form the analytical basis for estimating changes in energy efficiency, component condition, and emission-related performance [34].

Therefore, the trained ML–AI models act as a common inference backbone for all virtual sensing outputs, enabling the consistent derivation of multiple metrics from a shared feature space (Figure 4).

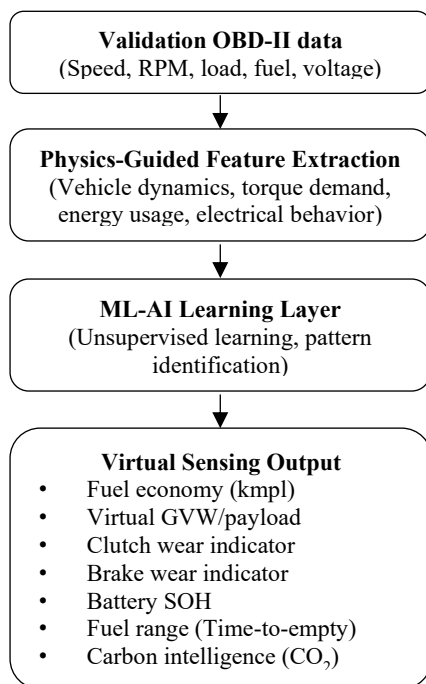


Figure 3. Virtual sensing.

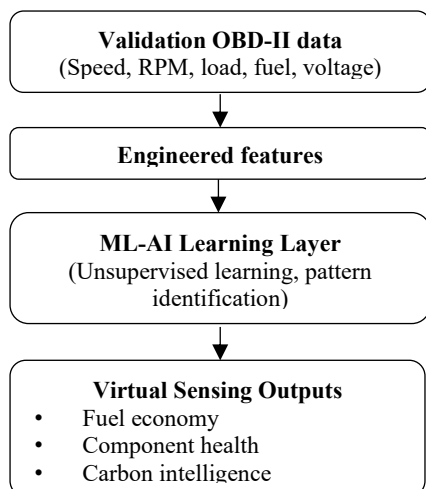


Figure 4. ML–AI models.

This unified modeling strategy improves robustness and scalability compared to single-metric or rule-based approaches and ensures coherence between the fuel economy, health, and carbon-related indicators. The formulation and application of individual virtual sensing metrics based on the learned representation are presented in Section 3.4.

Development of the Virtual Sensing Metrics (Final Version)

This section explains how the trained ML–AI models were used to derive individual virtual sensing metrics from real-world OBD-II data. Each metric is estimated using simple physical cause–effect relationships supported by the behavioral patterns learned from historical operating data. The ML–AI model learns the normal vehicle response characteristics and interprets deviations from these responses as changes in efficiency, component condition, fuel integrity, or emissions.

Virtual Fuel Economy

Fuel economy (FE) is estimated by analyzing the relationship between fuel consumption signals and

the distance traveled by vehicles over time. The ML–AI model learns typical fuel usage patterns under different operating conditions, such as steady cruising, acceleration, and stop–go traffic conditions. By correlating the fuel flow or fuel rate with the vehicle speed and distance, the model estimates the real-world FE in kilometers per liter (kmpl). This approach captures the influence of transient driving behavior, traffic congestion, and load demand, resulting in an FE estimate that reflects the actual in-use vehicle performance rather than the idealized laboratory conditions.

Virtual Clutch Wear Indicator

Clutch wear was inferred by analyzing the relationship between the engine speed, vehicle speed, and load demand. Under healthy clutch conditions, a stable and predictable relationship exists between the engine revolutions per minute (RPM) and vehicle speed for a given load. The ML–AI model learns this normal behavior during healthy operations. As clutch wear progresses, a higher engine speed is required to achieve the same vehicle speed, particularly during acceleration and load changes. A persistent deviation from the learned RPM–speed relationship, wherein the engine speed increases without a proportional increase in the vehicle speed, is interpreted as a clutch slip. The accumulation of slip events over time is used to estimate clutch wear trends and provide an early warning of degradation.

Virtual Brake Wear Indicator

Brake wear was estimated based on the effectiveness of the braking events using the vehicle speed response. During braking, a direct relationship exists between the brake application and the resulting reduction in the vehicle speed. The ML–AI model learns the typical deceleration behavior for given braking conditions and the operating states. As the brake components wear, a greater braking effort is required to achieve the same level of speed reduction. This appears as a reduced deceleration or a longer stopping response for similar braking events over time. By monitoring repeated braking events and comparing the observed speed reduction with the learned healthy behavior, the model identifies a gradual loss of braking effectiveness and estimates brake wear trends, thereby enabling condition-based maintenance of the brake system.

Virtual Battery State-of-Health (SOH)

The battery state-of-health was estimated using the electrical system behavior observed through the OBD-II voltage signals. The model learns the normal voltage response patterns during engine cranking, idle operation, and varying electrical loads. As the battery degrades, an increased voltage drop during cranking and slower voltage recovery are observed. By comparing the current voltage behavior with the learned healthy patterns, the model estimates battery degradation trends, thereby enabling the early identification of battery health issues without dedicated battery monitoring hardware.

Virtual Fuel Range and Fuel Integrity Monitoring

The remaining fuel range, expressed as time-to-empty or distance-to-empty, is estimated by combining virtual FE with fuel level trends and recent driving behavior. Instead of relying on fixed average consumption values, the ML–AI model adapts the range estimate based on current traffic conditions, driving intensity, and fuel usage patterns. Meanwhile, fuel level behavior is continuously monitored to assess fuel integrity. Under normal operation, the fuel level decreases gradually in proportion to engine operation and distance traveled. Sudden drops in fuel level when the vehicle is stationary or operating under low-consumption conditions are identified as abnormal. By correlating fuel level changes with vehicle speed, ignition status, and operating context, the model can flag potential fuel theft or leakage events using the same OBD-II data source.

Virtual Carbon Intelligence

Carbon intelligence is derived directly from fuel consumption behavior using established fuel-to-emission conversion principles. Because carbon dioxide (CO₂) emissions are proportional to the fuel burned, the virtual fuel consumption estimated by the ML–AI model was multiplied by the standard emission factors corresponding to the fuel type.

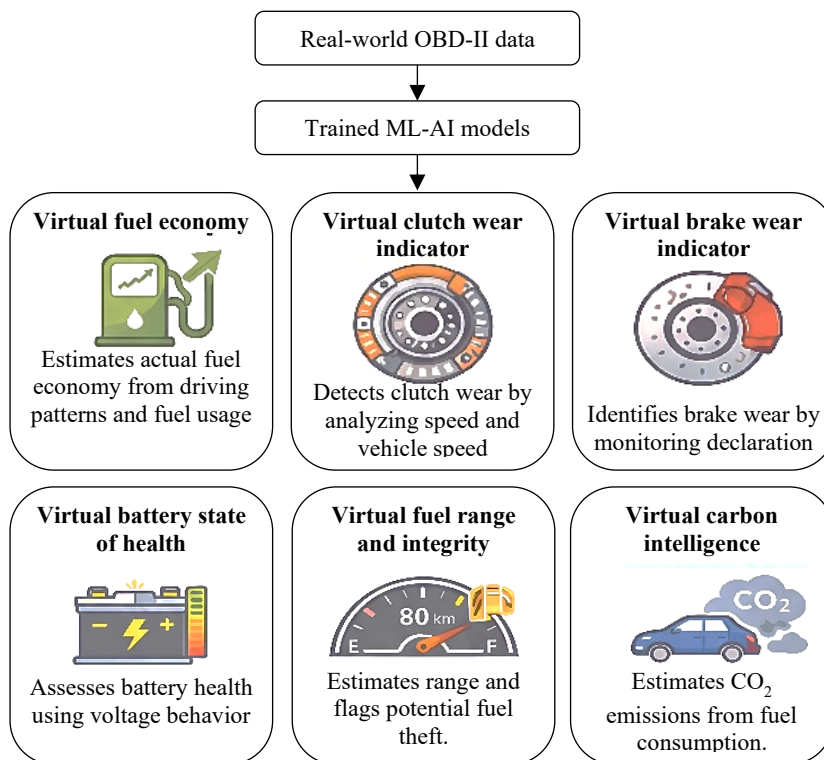


Figure 5. Virtual sensing metrics.

By continuously tracking fuel usage under real-world driving conditions, the model provides time-resolved and trip-level carbon-emission estimates. This enables the practical monitoring of vehicle carbon footprints and supports data-driven sustainability assessments without the need for exhaust gas sensors or laboratory-based emission testing (Figure 5).

RESULTS AND THE DISCUSSION

Overall, the proposed virtual sensing approach demonstrated the ability to extract multiple energy, health, and sustainability-related indicators from standard OBD-II data using a unified ML–AI framework. By translating real-world vehicle behavior into interpretable metrics, this approach supports realistic FE assessments, early identification of component degradation, reliable fuel range estimation with fuel integrity monitoring, and continuous carbon emission tracking. These outcomes highlight the practical value of virtual sensing for improving operational efficiency, reducing the total cost of ownership, and supporting data-driven sustainability initiatives without additional hardware.

The Virtual Fuel Economy Results

This subsection analyzes the real-world FE behavior derived from OBD-II data and interpreted using ML-based clustering. The objective was to examine how FE varies with vehicle speed under naturalistic driving conditions and how the ML model organizes this behavior into distinct and interpretable operating regimes.

Figure 6 shows the relationship between the vehicle speed and FE, where the ML model groups the operating points into three distinct FE clusters. The blue cluster (C1—low FE) represents a low fuel economy operation, which is predominantly observed under congested and stop-and-go driving conditions. In this region, frequent acceleration–deceleration cycles, idling, and inefficient engine operating points result in poor FE, typically below approximately 11 kmpl.

The green cluster (C2—moderate FE) corresponds to moderate fuel-economy behavior and is associated with smooth driving at moderate speeds. This operating regime reflects reduced transient

activity and a more stable throttle application, leading to improved but not optimal fuel efficiency, typically in the range of ~12–15 kmpl.

The orange cluster (C3—high or good FE) represents high fuel economy operation and is predominantly observed at steady cruising speeds under free-flow conditions. In this regime, a stable vehicle speed, minimal braking, and efficient engine load utilization enable an optimal FE, typically above ~16 (16–19) kmpl. The clear vertical separation between the clusters confirms that the ML approach successfully captures meaningful real-world FE regimes rather than arbitrary groupings.

Overall, this result demonstrates that the ML–AI-based virtual sensing approach can effectively structure complex real-world driving data into physically interpretable FE zones. Such clustering provides a strong foundation for behavior-aware efficiency analysis, driver guidance, and downstream virtual sensing applications related to maintenance and sustainability (Table 1).

Virtual Sensing Implementation and the Resulting Implications

Overall, the proposed method demonstrates an AI-enabled virtual sensing approach that leverages real-world OBD-II big data to estimate FE, component health, fuel integrity, and carbon intelligence without the need for additional physical sensors.

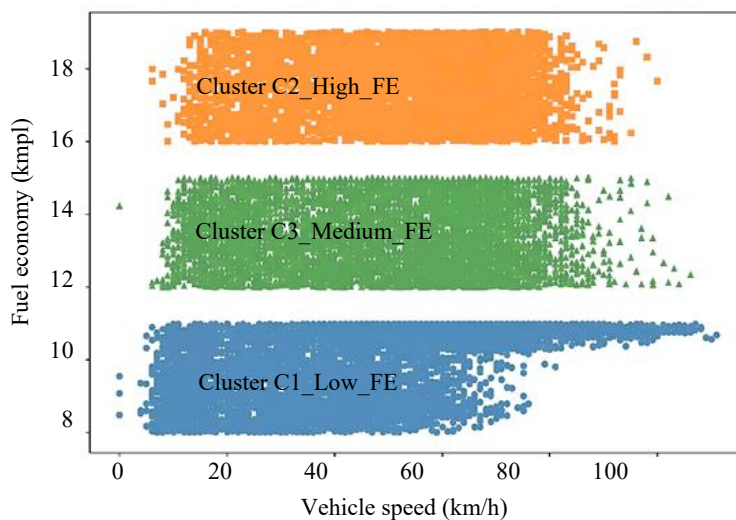


Figure 6. Fuel economy–speed relationship across ML-identified driving behavior clusters.

Table 1. Virtual sensing metric.

Virtual sensing metric	Estimation logic (brief)	Practical benefit
Fuel economy (kmpl)	Learns fuel usage per distance under real driving using OBD-II fuel and speed data	Realistic fuel efficiency assessment and driving optimization
Clutch wear indicator	Identifies a persistent mismatch between engine speed and vehicle speed under load	Early detection of clutch degradation and reduced breakdown risk
Brake wear indicator	Evaluates the reduction in vehicle speed for similar braking events over time	Predictive brake maintenance and improved braking safety
Battery health	Analyses voltage drops and recovery behavior during cranking and load changes	Prevention of unexpected battery failures
Fuel range and integrity	Combines fuel level trends with consumption patterns; flags sudden abnormal drops	Accurate range estimation and fuel theft or leakage detection
Carbon intelligence (CO ₂)	Converts real-world fuel consumption into emissions using standard emission factors	Continuous carbon footprint monitoring and environmental, social, and governance (ESG) reporting support

By combining physics-guided feature extraction with ML–AI models, this approach infers multiple interrelated vehicle performance and sustainability metrics under actual driving conditions. The ability to detect component degradation in advance enables a shift from scheduled maintenance to condition-based maintenance, thereby reducing unnecessary service actions, maintenance costs, and vehicle downtime. In parallel, adaptive fuel range estimation and fuel integrity monitoring improve operational reliability, whereas continuous carbon intelligence directly supports the reduction of CO₂ and greenhouse gas emissions. Overall, the proposed method demonstrates how software-defined vehicle intelligence can simultaneously enhance efficiency, lower the total cost of ownership, and contribute to sustainable and low-carbon mobility.

CONCLUSION

This study demonstrated an ML–AI-enabled virtual sensing framework that leverages real-world OBD-II big data to estimate the FE, component health indicators, fuel integrity, and carbon intelligence without the use of additional physical sensors. By integrating physics-guided feature extraction with unsupervised learning, multiple interrelated performance and sustainability metrics were reliably inferred under real-world driving conditions. The results highlight the feasibility of shifting from sensor-heavy and schedule-based approaches to scalable, software-defined vehicle intelligence that supports fuel efficiency improvement, condition-based maintenance, and real-world carbon monitoring. Overall, the proposed approach contributes to reduced fuel consumption, lower emissions, and improved total cost of ownership, offering a practical pathway toward intelligent and sustainable vehicle operations.

Nomenclature

AI	artificial intelligence
ML	machine learning
BDA	big data analytics
OBD-II	on-board diagnostics II
FE	fuel economy
kmpl	kilometers per liter
SOH	state-of-health
GVW	gross vehicle weight
CO ₂	carbon dioxide
GHG	greenhouse gas
RPM	revolutions per minute
TTE	time-to-empty
ECU	engine control unit
CAN	controller area network
ML–AI	machine learning and artificial intelligence
ESG	environmental, social, and governance

Acknowledgment

The author expresses sincere gratitude to Dr. R. Rajavel, Dr. S. Karthikeyan, Dr. M. Pugazhvadivu, and Dr. A. Suresh for their research guidance and consultation, as well as for their support with data collection, valuable feedback, and continuous encouragement throughout this research.

REFERENCES

1. Bishop CM. Pattern Recognition and Machine Learning. New York: Springer; 2006.
2. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer; 2009.
3. Géron A. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow. Sebastopol (CA): O'Reilly Media; 2019.
4. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. doi:10.1023/A:1010933404324.

5. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Advances in neural information processing systems*; 2017. p. 4765–4774.
6. Franco A, López R, Gallego J. OBD-based monitoring of vehicle energy use under real driving conditions. *IEEE Access*. 2019;7:103211–103223.
7. Lin S, Ma T, Javed MS. Prefeasibility study of a distributed photovoltaic system with pumped hydro storage for residential buildings. *Energy Convers Manag*. 2020;222:113199. doi:10.1016/j.enconman.2020.113199.
8. Ma S, Zhao X, Liu Y. Real-world driving cycle construction using large-scale OBD data. *Transp Res Part D*. 2020;87:102498.
9. Wu Y, Song J, Yu H. Big data driven analysis of vehicle energy consumption. *Appl Energy*. 2020;275:115410.
10. SAE International. Use of OBD data for vehicle performance and fuel economy analysis. *SAE Tech Pap*. 2018;2018-01-0399.
11. Kadlec P, Gabrys B, Strandt S. Data-driven soft sensors in the process industry. *Comput Chem Eng*. 2009;33(4):795–814. doi:10.1016/j.compchemeng.2008.12.012.
12. Fortuna L, Graziani S, Rizzo A, Xibilia MG. *Soft Sensors for Monitoring and Control of Industrial Processes*. London: Springer; 2007.
13. Ge Z, Song Z, Ding SX, Huang B. Data mining and analytics in the process industry: the role of machine learning. *IEEE Access*. 2017;5:20590–20616. doi:10.1109/ACCESS.2017.2756872.
14. Wang L, Wang Y, Zhang L. Virtual sensing for automotive applications: concepts and challenges. *IEEE Trans Veh Technol*. 2019;68(11):10488–10500.
15. Jardine AKS, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech Syst Signal Process*. 2006;20:1483–1510. doi:10.1016/j.ymssp.2005.09.012.
16. Lei Y, Li N, Guo L, Li N, Yan T, Lin J. Machinery health prognostics: a systematic review from data acquisition to RUL prediction. *Mech Syst Signal Process*. 2018;104:799–834. doi:10.1016/j.ymssp.2017.11.016.
17. Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX. Deep learning and its applications to machine health monitoring. *Mech Syst Signal Process*. 2019;115:213–237. doi:10.1016/j.ymssp.2018.05.050.
18. Hannan A, Arebey M, Begum R. Machine learning based predictive analytics for automotive applications. *IEEE Access*. 2020;8:18187–18204.
19. Rakha H, Ahn K, Trani A. Development of VT-Micro model for estimating hot stabilized light duty vehicle and truck emissions. *Transp Res Part D*. 2004;9(1):49–74. doi:10.1016/S1361-9209(03)00054-3.
20. Feng Y, Liu Y, Cao Y, Gong K, Liu S, Qin J. Thermal management evaluation for advanced aero-engines using catalytic steam reforming of hydrocarbon fuels. *Energy*. 2020;193:116738. doi:10.1016/j.energy.2019.116738.
21. Chen Y, Liu S, Li L. Fuel consumption estimation based on real-world driving data using machine learning. *Appl Energy*. 2019;238:1268–1280.
22. Fontaras G, Samaras Z. On-road fuel consumption and CO₂ emissions of light-duty vehicles. *Energy Policy*. 2008;36(10):3579–3588.
23. Barth M, Boriboonsomsin K. Real-world carbon dioxide impacts of traffic congestion. *Transp Res Rec*. 2008;2058:163–171. doi:10.3141/2058-20.
24. Zakaria Z, Kamarudin SK, Abd Wahid KA, Abu Hassan SH. The progress of fuel cell for Malaysian residential consumption: energy status and prospects for introduction as a renewable power generation system. *Renew Sustain Energy Rev*. 2021;144:110984. doi:10.1016/j.rser.2021.110984.
25. Zhao S, Li M, Wang H. Artificial intelligence for sustainable transportation systems. *J Clean Prod*. 2019;234:114–125.
26. Tao F, Qi Q, Wang L, Nee AYC. Digital twins and cyber-physical systems toward smart manufacturing and Industry 4.0: correlation and comparison. *Engineering*. 2019;5(4):653–661. doi:10.1016/j.eng.2019.01.014.

-
27. Boschert S, Rosen R. Digital Twin – the Simulation Aspect. In: Hehenberger P, Bradley D, editors. *Mechatronic Futures*. Cham: Springer; 2016. p. 59–74. doi:10.1007/978-3-319-32156-1_5.
 28. Zhu L, Yu FR, Wang Y, Ning B, Tang T. Big data analytics in intelligent transportation systems: a survey. *IEEE Trans Intell Transp Syst*. 2019;20(1):383–398. doi:10.1109/TITS.2018.2815678.
 29. Li Y, Zhang X, Wang Z. Deep learning and machine learning for vehicle energy consumption prediction. *Appl Energy*. 2020;257:113987.
 30. Meseguer JE, Puig V, Quevedo J. Driving styles and fuel consumption characterization using data-driven methods. *IEEE Intell Transp Syst Mag*. 2018;10(1):50–62.
 31. Qi Y, Luo L, Wang S. Driving behavior analysis based on vehicle trajectory big data. *Transp Res Part C*. 2018;95:234–247.
 32. Hu X, Yuan H, Zou C, Li Z, Zhang L. Battery state-of-health estimation using machine learning. *IEEE Trans Transp Electrific*. 2018;4(4):111–120.
 33. He W, Williard N, Osterman M, Pecht M. Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method. *J Power Sources*. 2011;196:10314–10321. doi:10.1016/j.jpowsour.2011.08.040.
 34. Jain A, Aslam M, Mehmood N. Machine learning applications in the automotive industry: a comprehensive review. *IEEE Access*. 2021;9:10112–10130.
 35. Chen M, Xu S, Zhao Y. AI-based energy modeling for intelligent vehicles. *Energy AI*. 2021;6:100095.