STM JOURNALS

Journal of
Automobile Engineering
and Applications

ISSN: 2455-3360
Volume 13, Issue 1, 2026
January—April
DOI Journal: 10.37591/JoAEA

https://journals.stmjournals.com/joaea

Research JoAEA

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

Arumugam Palanichamy[1,*], Karthikeyan Subramanian[2],
Rajavel Rangasamy[3], Suresh Alex Selvaraj[4]

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 onboard 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 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

[1]Research Scholar, Department of Marine Engineering, AMET University, Chennai, Tamil Nadu, India
[2]Assistant General Manager, Department of Fuel Cell Development, Ashok Leyland Technical Centre, Chennai, Tamil Nadu, India
[3]Principal, Head of Engineering Department, Sri Balaji Chockalingam Engineering College, Irumbedu, Arni, Tamil Nadu, India
[4]Professor, 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): 39–50p.

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 On-Board Diagnostics II (OBD-II) data generated during real-world vehicle operation. In this study, such data are 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 as interconnected aspects of real-world vehicle operation [6–9].

Traditional vehicle monitoring and maintenance approaches in current production vehicles rely heavily on physical sensors, scheduled inspections, and rule-based thresholds, which increase system cost, complexity, and maintenance overhead [10–15]. In this context, such 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 extract multiple forms of vehicle intelligence directly 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, enabling 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 within OBD-II data.

Virtual sensing, also referred to 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]. While 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 integrated sensing of energy, health, and emissions [14, 29].

In parallel, accurate assessment of real-world fuel economy and carbon emissions has become increasingly important for regulatory compliance, sustainability assessment, 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 operation, 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 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 – are derived from a unified data source.

The key contributions of this work 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 original equipment manufacturers and fleet operators to improve operational efficiency, reduce fuel consumption and emissions, and enhance vehicle intelligence using existing vehicle data infrastructure.

PROBLEM STATEMENT OVERVIEW

The increasing deployment of connected vehicles has enabled continuous acquisition of rich operational data through OBD-II systems under real-world driving conditions [6, 8, 9]. However, most existing research efforts and practical implementations utilize this 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 that is 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 machine learning and artificial intelligence 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 are largely 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 adopts a data-driven methodology that integrates big data analytics, machine learning, and artificial intelligence 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 reliance on additional physical sensors. Such an approach aligns with recent trends in automotive analytics that emphasize scalable, software-defined intelligence built on existing onboard data infrastructure [6, 9, 28].

Real-world vehicle data acquired through OBD-II systems exhibit 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 for analyzing fuel consumption behavior [7, 20, 21], driving patterns [30, 31], and emissions under real-world conditions [22, 23]. However, these efforts largely focus on individual performance metrics, motivating the need for an integrated methodology capable of extracting multiple forms of vehicle intelligence from a common data source.

To address this need, the proposed methodology is structured around the concept of virtual sensing, also referred to as soft sensing, where 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 due to their cost-effectiveness and robustness [11, 12], but their systematic application to automotive systems for simultaneous estimation of energy performance, component health, and carbon emissions remains limited [14, 29].

Machine learning and artificial intelligence 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, where 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 interpretability and physical consistency of the derived virtual metrics, which is 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 onboard 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 in the upcoming sections (Figure 1).

OBD-II Data Acquisition and Big Data Characterization Process

Real-world vehicle operational data were collected through the On-Board Diagnostics II (OBD-II) interface during 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, enabling 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.

The acquired OBD-II dataset exhibits essential big data characteristics, including high volume and velocity due 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 removal of communication dropouts, filtering of physically inconsistent values, and consistency checks based on 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 operation and forms the foundation for subsequent physics-guided feature extraction and ML–AI-based virtual sensing. 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 used to estimate variables that are not directly measured by physical sensors by exploiting existing signals and underlying physical relationships. In automotive applications, virtual sensing enables the estimation of performance, health, and sustainability-related metrics using standard onboard data, thereby reducing dependence on additional hardware sensors and improving scalability across vehicle platforms [11–14]. The continuous availability of OBD-II data during real-world vehicle operation makes this approach particularly suitable for developing software-defined vehicle intelligence.

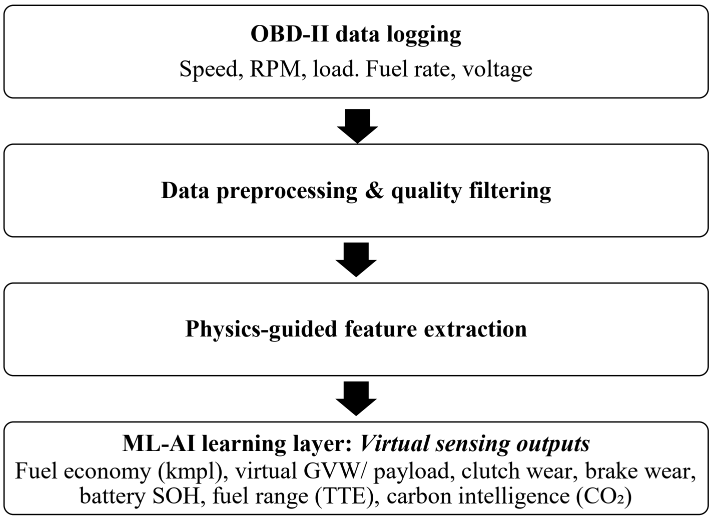

Figure 1. Methodology overview.

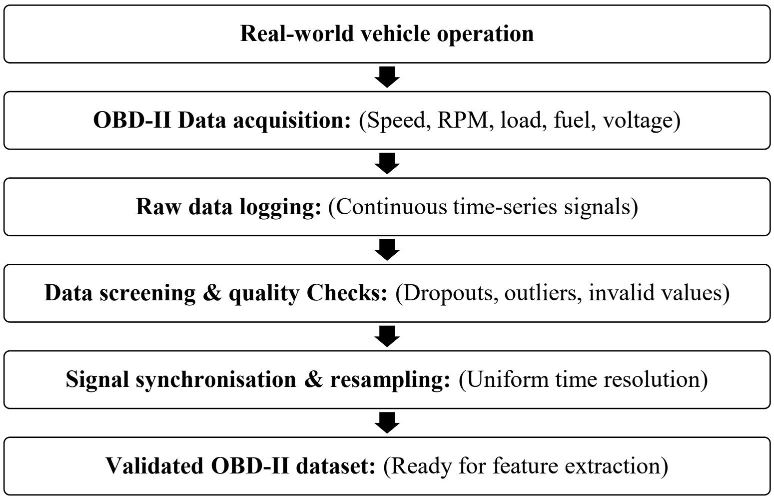

Figure 2. OBD-II Data processing.

The proposed system architecture follows a layered virtual sensing concept, where 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, ensuring interpretability and consistency with fundamental vehicle operating principles [3, 5, 14].

Machine learning and artificial intelligence techniques are then used to analyze the extracted features and estimate virtual sensing outputs. Unsupervised learning methods are particularly suitable for real-world vehicle data, as labeled fault or degradation data are rarely available and operating conditions vary continuously during normal driving [16, 17]. By learning typical operating behavior directly from data, the ML–AI layer supports 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 derived from a shared set of features. Using a common feature space allows these outputs to be estimated together, reflecting their natural interdependence during real-world vehicle operation (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 virtual sensing of fuel economy, component health, and carbon-related metrics. The emphasis is on learning normal vehicle operating behavior directly from real-world data to enable reliable sensor-less inference.

The engineered features are used as inputs to the ML–AI learning layer. Given the limited availability of labeled fault or degradation data in real-world vehicle operation, unsupervised learning techniques are employed to learn 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 operation across multiple dimensions. Deviations from this learned behavior form the analytical basis for estimating changes in energy efficiency, component condition, and emission-related performance [34].

The trained ML–AI models therefore act as a common inference backbone for all virtual sensing outputs, enabling multiple metrics to be derived consistently from a shared feature space (Figure 4).

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

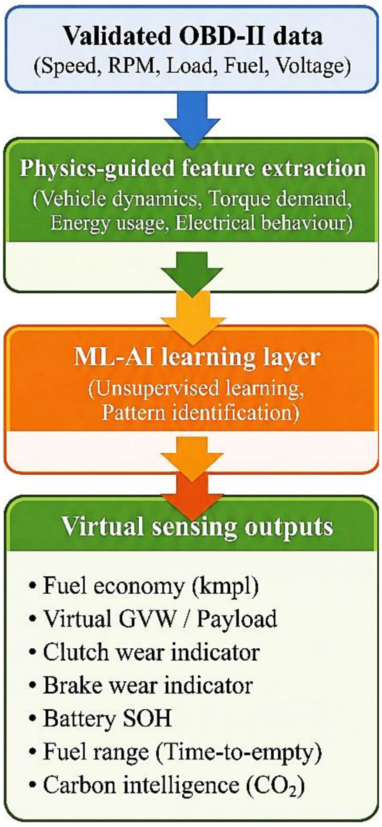

Figure 3. Virtual sensing.

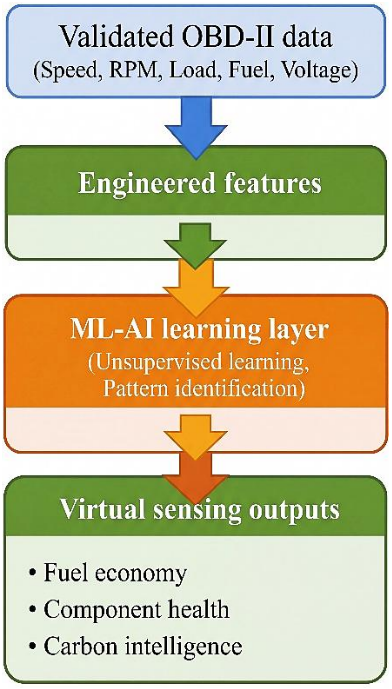

Figure 4. ML – AI models.

Development of the Virtual Sensing Metrics (Final Version)

This section explains how the trained ML–AI models are 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 behavioral patterns learned from historical operating data. The ML–AI model learns normal vehicle response characteristics and interprets deviations from these responses as changes in efficiency, component condition, fuel integrity, or emissions.

Virtual Fuel Economy (FE)

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

Virtual Clutch Wear Indicator

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

Virtual Brake Wear Indicator

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

Virtual Battery State of Health (SOH)

Battery state of health is estimated using electrical system behavior observed through OBD-II voltage signals. The ML–AI model learns normal voltage response patterns during engine cranking, idle operation, and varying electrical loads. As the battery degrades, increased voltage drop during cranking and slower voltage recovery are observed. By comparing current voltage behavior with learned healthy patterns, the model estimates battery degradation trends, allowing 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 fuel economy 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. In parallel, fuel level behavior is continuously monitored to assess fuel integrity. Under normal operation, 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. Since carbon dioxide (CO_2) emissions are proportional to fuel burned, the virtual fuel consumption estimated by the ML–AI model is multiplied by standard emission factors corresponding to the fuel type. By continuously tracking fuel usage under real-world driving conditions, the model provides time-resolved and trip-level carbon emission estimates. This enables practical monitoring of vehicle carbon footprint and supports data-driven sustainability assessment without the need for exhaust gas sensors or laboratory-based emission testing (Figure 5).

RESULTS AND THE DISCUSSION

Overall, the proposed virtual sensing approach demonstrates 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, the approach supports realistic fuel economy assessment, 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 total cost of ownership, and supporting data-driven sustainability initiatives without additional hardware sensors.

The Virtual Fuel Economy Results

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

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

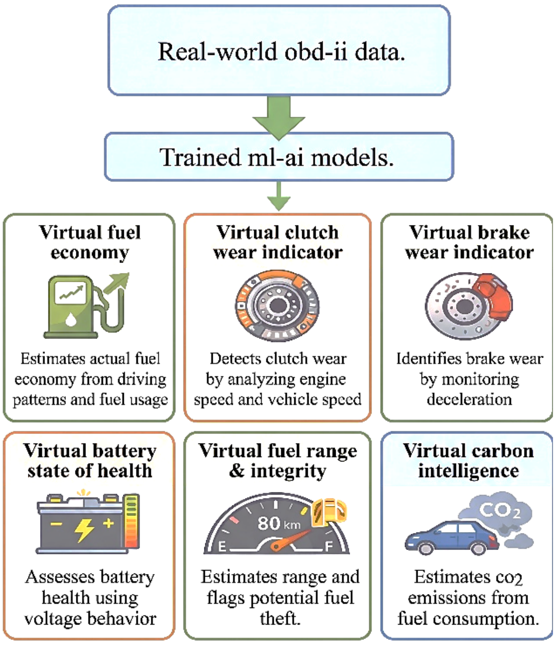

Figure 5. Virtual sensing metrics.

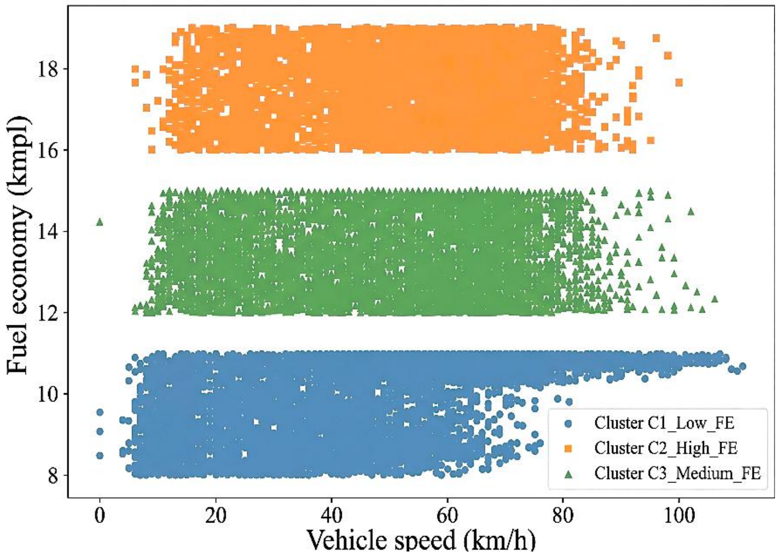

Figure 6. Fuel economy–speed relationship across ML-identified driving behavior clusters.

The green cluster (C2 – Moderate FE) corresponds to moderate fuel economy behavior and is associated with smoother driving at moderate speeds. This operating regime reflects reduced transient activity and 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 traffic conditions. In this regime, stable vehicle speed, minimal braking, and efficient engine load utilization enable optimal fuel economy, typically above ~16 (16–19) kmpl. The clear vertical separation between clusters confirms that the ML approach successfully captures meaningful real-world fuel economy regimes rather than arbitrary groupings.

Overall, this result demonstrates that the ML–AI-based virtual sensing approach effectively structures complex real-world driving data into physically interpretable fuel economy 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).

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 optimisation
Clutch wear indicator	Identifies persistent mismatch between engine speed and vehicle speed under load	Early detection of clutch degradation and reduced breakdown risk
Brake wear indicator	Evaluates reduction in vehicle speed for similar braking events over time	Predictive brake maintenance and improved braking safety
Battery Health	Analyses voltage drop and recovery behaviour during cranking and load changes	Prevention of unexpected battery failures
Fuel range & 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 ESG reporting support

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 fuel economy, component health, fuel integrity, and carbon intelligence without additional physical sensors. By combining physics-guided feature extraction with ML–AI models, the 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, reducing unnecessary service actions, maintenance cost, and vehicle downtime. In parallel, adaptive fuel range estimation and fuel integrity monitoring improve operational reliability, while continuous carbon intelligence directly supports CO_2 and greenhouse gas emission reduction. Overall, the proposed method demonstrates how software-defined vehicle intelligence can simultaneously enhance efficiency, lower total cost of ownership, and contribute meaningfully to sustainable and low-carbon mobility.

RESEARCH CONCLUSION

This study demonstrated an ML–AI-enabled virtual sensing framework that leverages real-world OBD-II big data to estimate fuel economy, 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 inferred reliably under actual driving conditions. The results highlight the feasibility of shifting from sensor-heavy and schedule-based approaches toward 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 operation.

The Nomenclature

Table 2 depicts the nomenclature.

Table 2. Nomenclature.

Symbol / Term	Description
AI	Artificial intelligence
ML	Machine learning
BDA	Big data analytics
OBD-II	On-board diagnostics II
FE	Fuel economy
kmpl	Kilometres per litre
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, Dr A Suresh for their research guidance and consultation and extended their support for data collection, valuable feedback, and continuous encouragement throughout the course of this research work.

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. Machine Learning. 2001;45(1):5–32.
5. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765–4774.
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. Wang J, Liu H, Chen Y. Machine learning models for vehicle fuel consumption prediction using OBD data. Energy Convers Manag. 2020;222:113199.
8. Ma S, Zhao X, Liu Y. Real-world driving cycle construction using large-scale OBD data. Transp Res Part D Transp Environ. 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.
12. Fortuna L, Graziani S, 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. J Process Control. 2017;67:123–138.
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(7):1483–1510.
16. Lei Y, Li N, Guo L, Li N, Yan T, Lin J. Machinery health prognostics: A systematic review. Mech Syst Signal Process. 2018;104:799–834.
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.
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 fuel consumption. Transp Res Part D Transp Environ. 2004;9(1):49–74.
20. Sun M, Chen X, Zhang Y. Real-world vehicle fuel consumption modeling based on big data analytics. Energy. 2020;193:116738.
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 CO_2 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.
24. Hasan MK, Feng S. A review of data-driven energy consumption models for vehicles. Renew Sustain Energy Rev. 2021;144:110984.
25. Zhao S, Li M, Wang H. Artificial intelligence for sustainable transportation systems. J Clean Prod. 2019;234:114–125.
26. Tao F, Zhang H, Liu A, Nee AYC. Digital twins and cyber–physical systems toward smart manufacturing. Engineering. 2019;5(4):653–661.

27. Boschert S, Rosen R. Digital twin—The simulation aspect. In: Mechatronic futures. Cham: Springer; 2016. p. 59–74.
28. Zhang J, Wang Y, Li L. Big data analytics in intelligent transportation systems. IEEE Trans Intell Transp Syst. 2019;20(9):3436–3448.
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 Emerg Technol. 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 Electrification. 2018;4(4):111–120.
33. He W, Williard N, Chen C, Pecht M. State of health estimation of lithium-ion batteries. J Power Sources. 2011;196(23):10314–10321.
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.