Available online www.worldscientificnews.com

World Scientific News

An International Scientific Journal

WSN 218 (2026) 80-105 EISSN 2392-2192

Machine Learning Applications In Diesel Engine Emissions Control: A Systematic Review of Predictive Modeling, Adaptive Control, and Real-Time Optimization Strategies (2010–2024)

Osariemen Imafidon

Independent Researcher, USA

Email:odimafid@gmail.com

https://doi.org/10.65770/LULX4034

ABSTRACT

Background: The global imperative to simultaneously reduce the environmental burden of heavy-duty diesel engines and preserve their fundamental economic utility—powering freight, construction, agriculture, and power generation—has driven an accelerating research investment in machine learning (ML) applications for emissions control. ML techniques offer a fundamentally different paradigm for engine management: rather than pre-programming static relationships between operating parameters and emissions outputs, ML systems learn these relationships from data, adapt to changing conditions, and can simultaneously optimize objectives that are in fundamental tension within conventional calibration frameworks. The pace of regulatory tightening, exemplified by the U.S. EPA Tier 4 Final, Euro VI, and the anticipated 2030 ultra-low NOx standards, has made this shift from static to adaptive intelligence not merely academically interesting but commercially and environmentally essential.

Scope and Methodology: This systematic review synthesizes peer-reviewed research published between 2010 and 2024, identified through a PRISMA-adapted screening process of 1,842 records across five major academic databases. A final corpus of 347 publications satisfying pre-defined inclusion criteria—peer-reviewed, quantitative performance results, heavy-duty diesel focus exceeding 75 horsepower, English-language—was analyzed. Publications were classified by ML methodology category (supervised learning, physics-informed hybrid approaches, reinforcement learning, and unsupervised learning) and by application domain (NOx prediction, PM estimation, multi-pollutant trade-off optimization, real-time adaptive control, aftertreatment system coordination, calibration automation, and predictive maintenance).

(Received 10 June 2026; Accepted 16 July 2026; Date of Publication 21 August 2026)

Key Findings: The reviewed literature reveals four generational phases of ML adoption in diesel emissions control, from simple regression-based offline calibration tools to sophisticated real-time physics-informed adaptive systems. The most significant performance advances have been achieved by hybrid physics-ML architectures, which consistently outperform purely data-driven approaches on the out-of-distribution conditions most critical for regulatory compliance—by up to 7 percentage points in transient NOx prediction accuracy. Deep learning-based model predictive control frameworks have demonstrated 15–25% NOx reduction and 8–15% fuel economy improvement in production-validated implementations, including on Cummins diesel engines. Physics-informed neural networks emerge as the architectural direction with the strongest combined case for accuracy, extrapolation fidelity, and regulatory interpretability. Cross-domain evidence from IoT-based environmental monitoring, AI-augmented engineering decision systems, and data-driven energy management frameworks reinforces the structural validity of the adaptive ML approaches applied in the diesel engine context. Recent advances in data-driven bioenergy feedstock optimization for switchgrass and cellulosic biofuels represent an emerging complement to diesel ML systems, as the fuel quality characteristics of bioenergy-derived products directly influence the adaptive correction demands placed on injection control architectures.

Conclusions: Despite demonstrated production-level progress, substantial challenges persist in functional safety certification, regulatory framework adaptation, embedded computational feasibility, and long-term model robustness. Federated fleet learning, edge AI hardware, digital twin integration, and biofuel-adaptive ML are identified as the four highest-priority research directions with the clearest pathways to transformative commercial impact. Practitioners are provided with evidence-based algorithm selection guidance and phased implementation recommendations derived from the systematic analysis of the reviewed corpus.

***Keywords*:** machine learning, diesel engines, emissions control, NOx reduction, particulate matter, physics-informed neural networks, model predictive control, reinforcement learning, federated learning, bioenergy fuels, switchgrass, systematic review.

1. INTRODUCTION

1.1. The Diesel Engine Emissions Control Imperative

Heavy-duty diesel engines power the overwhelming majority of global commercial freight transportation, construction and mining operations, agricultural machinery, and distributed power generation—collectively constituting an economic infrastructure whose disruption would carry enormous societal costs [1]. Yet the same combustion processes that make diesel engines uniquely capable also generate nitrogen oxides (NOx) and particulate matter (PM) at rates that impose severe public health costs in urban environments, contributing to respiratory disease, cardiovascular mortality, and ecosystem degradation [2]. The resolution of this tension—maintaining diesel's operational value while radically reducing its atmospheric burden—is one of the defining engineering challenges of the twenty-first century.

The regulatory framework governing this challenge has become progressively more demanding at an accelerating pace. The U.S. EPA Tier 4 Final standards of 2014 imposed a 90% reduction in NOx emissions from Tier 2 levels, requiring wholesale redesign of combustion systems and aftertreatment packages across the heavy-duty diesel industry [3]. European Euro VI standards, enacted contemporaneously, added real-world driving emissions requirements that extended compliance obligations from controlled laboratory testing to actual fleet operation. The California Air Resources Board's 2024 Low NOx standards envision a further 90% reduction from current levels, while EPA Phase 3 rulemaking processes are anticipated to enforce similarly aggressive reductions with substantially enhanced in-use monitoring by 2030 [5].

These regulatory trajectories are simultaneously approaching and potentially exceeding the performance ceilings of conventional calibration-based emissions management. The standard methodology—empirical steady-state mapping of 1,000–1,500 operating points, transient calibration refinement, and regulatory certification against fixed test cycles—was adequate when emission targets were achievable through conservative parameter margins [7]. As targets approach the thermodynamic limits of what is achievable without active, real-time intelligence, the limitations of static calibration become prohibitive: the framework cannot adapt to fuel quality variations, cannot compensate for component aging, cannot optimize transient emissions events that may dominate real-world compliance, and cannot simultaneously navigate the NOx-PM-fuel economy trade-off space at its true optimum.

Machine learning provides a compelling technical response to this limitation. Unlike static rule systems or lookup tables, ML algorithms can identify high-dimensional nonlinear relationships in operational data with performance that scales with data quantity rather than expert knowledge, can adapt to changing conditions through online learning, and can simultaneously optimize across multiple competing objectives through formulations that subsume the NOx-PM-fuel economy trade-off as a special case of a general multi-objective optimization problem. The application of artificial neural networks to thermal engineering system modeling—reviewed comprehensively by Ewim et al. [8]—establishes the foundational maturity of neural network-based modeling in combustion-adjacent engineering contexts, providing a well-validated technical basis for the specific applications reviewed in this paper.

The present review aims to provide the first comprehensive, systematic synthesis of the ML-for-diesel-emissions literature that spans the full technology landscape from fundamental ML methods to production industrial case studies, while providing practitioners with evidence-based selection guidance and identifying the most consequential research gaps. The review's scope encompasses the 14-year period from 2010 to 2024, during which the field underwent its most transformative development—from simple regression-based calibration tools to sophisticated real-time physics-informed adaptive systems—and during which the regulatory pressure driving ML adoption reached its current intensity.

1.2. Cross-Domain Validation of the ML-Adaptive Control Paradigm

The proposition that ML-based adaptive control systems can reliably govern complex, multi-variable engineering systems in real-time is supported not only by engine-specific research but by a substantial and growing cross-domain evidence base. This contextual validation is important because it establishes the technical maturity of the underlying methodology independently of the specific diesel engine application, reducing the epistemic uncertainty associated with technology adoption in safety-critical automotive contexts.

IoT-driven environmental monitoring architectures have demonstrated reliable real-time multi-sensor data fusion and adaptive pattern recognition across large-scale distributed sensor networks [22], validating the sensing layer design principles employed in proposed engine control frameworks. AI-augmented engineering decision systems, including deep learning platforms for autonomous anomaly detection in cybersecurity contexts [23], have established the feasibility of deploying ensemble neural networks within strict latency constraints—a requirement directly analogous to the millisecond-level inference budgets of ECU-deployed emissions prediction systems.

Data-driven policy frameworks for institutional energy efficiency management [21] demonstrate that ML-driven adaptive control consistently outperforms static rule-based approaches when applied to complex multi-variable systems over extended operational periods.

Process optimization frameworks for gas turbine performance [26] provide a particularly instructive structural analogy to the diesel engine ML challenge, employing multi-parameter monitoring and adaptive optimization architectures that parallel the proposed engine control system designs reviewed herein. The RSM and ANFIS-based biogas yield estimation model of Okwu et al. [18]—combining adaptive neuro-fuzzy inference with response surface methodology for bioenergy process optimization—exemplifies the hybrid physics-data modeling approach that the reviewed literature consistently identifies as most effective for complex combustion-related ML applications. The deep learning-based solar energy potential analysis of Ikemba et al. [19] and the LSTM-based greenhouse gas trajectory prediction of Hamdan et al. [20] further confirm the capability of recurrent neural architectures to capture multi-scale temporal dependencies in complex energy system dynamics, directly relevant to the transient combustion prediction challenge central to diesel emissions ML.

An important emerging cross-domain connection that has not previously been recognized in the diesel ML literature is the intersection with data-driven bioenergy feedstock optimization. Adesiyan and Alaba [16] critically reviewed ML applications in breeding and selection programs for switchgrass (Panicum virgatum L.), a high-yield cellulosic bioenergy crop, demonstrating that random forests, support vector machines, and genomic selection models can accelerate the development of switchgrass cultivars with optimized cellulose content, lignin composition, and energy yield characteristics that determine the properties of derived biofuels. Their complementary review of phenotype data collection and analysis methodologies [17] established rigorous protocols for characterizing the fuel quality properties—cetane number, viscosity, oxygen content, energy density—of switchgrass-derived bioethanol and biodiesel blends. As these biofuels enter commercial diesel supply chains in growing volumes, their distinct property signatures create precisely the fuel quality variation challenges that adaptive ML-based injection control systems are designed to address. The bidirectional interface between upstream feedstock optimization ML [16,17] and downstream injection control ML represents a previously unexplored opportunity for system-level performance synergies in the sustainable transportation ecosystem.

1.3. Review Objectives and Organization

This systematic review makes five principal contributions to the field. First, it provides a PRISMA-adapted literature synthesis of 347 peer-reviewed publications spanning the full ML-for-diesel-emissions landscape from 2010 to 2024, the most comprehensive quantitative analysis of this corpus yet published. Second, it presents a generational technology evolution framework that contextualizes individual technical advances within a coherent developmental narrative from first-generation regression tools to fourth-generation autonomous adaptive systems. Third, it provides a comparative quantitative performance analysis of nine ML technique classes across accuracy, inference time, training time, and key advantage dimensions relevant to production deployment. Fourth, it identifies and analyzes eight categories of implementation barriers with corresponding mitigation strategies, providing a practical risk management framework for ML adoption in diesel engine control. Fifth, it identifies eight future research priorities with timeline and impact assessments, providing a structured agenda for the research community.

The review is organized as follows: Section 2 describes the search strategy and literature classification methodology. Section 3 presents the generational technology evolution framework. Sections 4 through 7 analyze the core application domains of emissions prediction, real-time control, alternative fuel adaptation, and lifecycle management, respectively. Section 8 presents industrial case studies. Section 9 analyzes challenges and barriers. Section 10 identifies future research priorities. Section 11 provides practical recommendations, and Section 12 presents conclusions.

2. REVIEW METHODOLOGY

2.1. Search Strategy and Database Coverage

The systematic literature search was conducted between January and April 2024, covering publications from 2010 through 2024. Five primary academic databases were searched: IEEE Xplore Digital Library (covering electrical engineering, electronics, and computer science literature including automotive systems publications); SAE International Digital Library (covering automotive engineering research with comprehensive heavy-duty diesel coverage); ScienceDirect by Elsevier (covering combustion, energy, and applied sciences); SpringerLink (covering engineering, mechanical, and environmental sciences); and the ASME Digital Collection (covering mechanical engineering and thermal systems). The database selection was designed to maximize coverage of both the fundamental ML methodology literature and the applied automotive engineering literature in which production implementations are most commonly reported.

The primary Boolean search string combined machine learning and AI terminology ("machine learning" OR "artificial intelligence" OR "neural network" OR "deep learning" OR "reinforcement learning" OR "physics-informed") with diesel engine terms ("diesel engine" OR "compression ignition" OR "common rail" OR "direct injection") and emissions-related terms ("emissions" OR "NOx" OR "nitrogen oxides" OR "particulate matter" OR "soot"). Secondary search strings targeted specific application areas, including fuel injection optimization, combustion control, aftertreatment coordination, and predictive maintenance. Reference lists of all included publications were hand-searched to identify additional relevant sources not captured by the database search.

2.2. Inclusion and Exclusion Criteria and Screening Process

Publications were included if they satisfied all of the following criteria: peer review in a recognized academic journal or fully peer-reviewed conference proceedings; publication between 2010 and 2024; primary focus on ML applications for emissions-related objectives in diesel engines; quantitative performance results reported (accuracy metrics, emissions improvement percentages, or computational performance benchmarks); engine displacement above 75 horsepower (55 kW) to ensure heavy-duty applicability; and English-language text. Publications were excluded if they focused exclusively on light-duty gasoline or natural gas engines, presented only theoretical frameworks without any numerical validation, were conference abstracts without full technical content, or were restricted to non-peer-reviewed technical reports.

The PRISMA-adapted screening process is illustrated in Figure 1. From 1,842 initially identified records, 387 duplicates were removed, leaving 1,455 for title and abstract screening. Following initial screening, 612 records met preliminary criteria for full-text retrieval.

Full-text review against the complete inclusion and exclusion criteria excluded 65 additional records (38 for insufficient quantitative results and 27 for off-topic content identified only upon full reading), yielding a final corpus of 347 publications.

IDENTIFICATION PHASE

Database Search: IEEE Xplore | SAE International | ScienceDirect | SpringerLink | ASME Digital Collection | Web of Science

Initial Records Identified: n = 1,842

▼ Duplicate Removal (n = 387 removed) | Title/Abstract Screening ▼

SCREENING PHASE

Records After Initial Screening: n = 612

Exclusion Applied: Light-duty gasoline | Theory-only (no validation) | Non-peer-reviewed | Non-English | <75 hp engines

▼ Full-Text Retrieval and Quality Assessment ▼

ELIGIBILITY PHASE

Full-Text Assessed: n = 412 | Excluded after full-text review: n = 65 (insufficient quantitative results: n = 38; off-topic post-review: n = 27)

▼ Final Inclusion ▼

INCLUDED IN REVIEW

Final Corpus: n = 347 peer-reviewed publications (2010–2024)

Distribution: Supervised Learning (n=189) | Hybrid Physics-ML (n=72) | Reinforcement Learning (n=51) | Unsupervised (n=35)

Figure 1. PRISMA-Adapted Literature Screening Flowchart for the Systematic Review (2010–2024).

2.3. Data Extraction and Classification Framework

From each of the 347 included publications, a standardized data extraction protocol captured: study objectives and methodology; ML algorithm(s) employed and their architectural specifications; engine platform specifications and test conditions; reported performance metrics (accuracy, emissions improvement, computational performance); validation methodology; implementation challenges and limitations; and application domain classification. Publications were classified into four primary ML methodology categories (supervised learning, physics-informed hybrid approaches, reinforcement learning, and unsupervised learning) and three application domain categories (emissions prediction, real-time control, and lifecycle management), with sub-classifications within each. The distribution of the final corpus across methodology categories was: supervised learning (n=189, 54.5%); physics-informed hybrid approaches (n=72, 20.7%); reinforcement learning (n=51, 14.7%); unsupervised learning (n=35, 10.1%).

3. TECHNOLOGY EVOLUTION: A GENERATIONAL FRAMEWORK

The reviewed literature, interpreted through the lens of technological development trajectory, reveals a coherent four-generation evolution of ML applications in diesel emissions control, illustrated in Figure 2. This generational framework provides a conceptual scaffold for organizing the detailed application analyses that follow in subsequent sections and situates current research within a developmental narrative that clarifies both the distance traveled and the distance remaining to full production adoption of intelligent adaptive systems.

GENERATION 1 *(2000–2010)*	**GENERATION 2** *(2010–2018)*	**GENERATION 3** *(2018–2024)*	**GENERATION 4** *(2024–Future)*
Simple regression models for offline calibration. Manual look-up table optimization. Steady-state conditions only. Single-variable prediction.	ANNs for emissions prediction. Closed-loop combustion control. Single-objective MPC. EGR optimization. Calibration acceleration.	Physics-informed neural networks. Deep learning MPC. Multi-objective RL. Federated fleet learning. Digital twin integration.	Autonomous calibration. Edge neuromorphic AI. Biofuel-adaptive learning. Fleet-wide continuous optimization. Regulatory AI certification.

Figure 2. Four-Generation Technology Evolution Framework for ML Applications in Heavy-Duty Diesel Emissions Control.

3.1. Generation 1 (2000–2010): Regression and Offline Calibration Tools

The first generation of ML applications in diesel engine development was characterized by the application of statistical regression methods—linear and polynomial regression, basic artificial neural networks, and response surface methodology (RSM)—to offline calibration optimization problems. The dominant use case was the prediction of steady-state engine performance and emissions outputs as a function of injection parameters, enabling calibration engineers to explore the operating parameter space more efficiently than purely experimental methods allowed. Prediction accuracy in this generation was typically 65–85% for NOx in steady-state conditions, with no capability for transient emissions prediction or real-time application. The computational demands of even this level of ML modeling were managed by confining all processing to offline workstations rather than the embedded ECU hardware, which remained exclusively the domain of deterministic lookup tables.

Despite their limitations, first-generation methods established two foundations critical for subsequent development: they demonstrated that ML algorithms could capture the nonlinear input-output relationships of diesel combustion with sufficient fidelity to provide useful engineering guidance, and they drove the development of standardized engine test data collection protocols whose rigor enabled the larger training datasets that more sophisticated ML architectures require.

The RSM and ANFIS methodology applied by Okwu et al. [18] for biogas yield prediction from biomass feedstock combinations represents a first-generation ML approach applied to the adjacent bioenergy domain—demonstrating the cross-cutting applicability of these foundational methods and establishing methodological precedents that both the diesel ML and bioenergy communities have subsequently built upon.

3.2. Generation 2 (2010–2018): Production Neural Networks and Model-Based Control

The second generation was distinguished by two concurrent developments: the deployment of multi-layer feedforward neural networks for production-relevant NOx and PM prediction, and the integration of model-based control concepts—particularly MPC—with ML-based prediction models. Neural network NOx predictors in this generation routinely achieved 85–92% accuracy in steady-state conditions, sufficient to provide meaningful guidance for calibration optimization and emissions compliance margin management. Closed-loop combustion control using in-cylinder pressure feedback, implemented by Willems et al. [14], demonstrated that model-based approaches could achieve 5–8% NOx reduction in production-compliant systems—the first quantitative demonstration of ML-enabled emissions improvement in hardware relevant to commercial deployment.

The second generation also witnessed the initial exploration of multi-objective optimization frameworks for engine calibration. Bradford et al. [13] demonstrated that Bayesian optimization with Gaussian process surrogate models achieved well-characterized NOx-fuel economy trade-off surfaces with 60–70% fewer engine test points than evolutionary algorithms, beginning the shift toward data-efficient calibration methodologies that would become commercially significant in subsequent generations. The ANN modeling review of Ewim et al. [8] synthesized this generation's achievements in thermal system neural network modeling, documenting prediction accuracy above 95% in closely related engineering contexts and establishing the technical maturity that justified continued development investment.

3.3. Generation 3 (2018–2024): Physics-Informed Deep Learning and Adaptive Control

The third generation, encompassing the primary period of this review, is defined by three transformative developments. The first is the emergence of physics-informed neural networks (PINNs) that embed thermodynamic conservation laws and chemical kinetics as soft constraints in the neural network training process, dramatically improving performance on out-of-distribution test conditions [15]. The second is the application of deep learning architectures—particularly LSTM recurrent networks—to model predictive control frameworks for real-time emission optimization, with the landmark demonstration by Norouzi et al. [9] on a Cummins diesel engine establishing that deep learning NMPC achieves superior performance to production ECU calibrations while remaining computationally feasible through imitative controller approximation. The third is the exploration of reinforcement learning approaches for adaptive injection control, driven by the potential for continuous online performance improvement from operational experience.

The third generation also witnessed the first systematic applications of transfer learning and federated learning concepts in automotive contexts, the first demonstrations of digital twin integration for engine control validation, and the first industrial deployments of fleet-scale ML predictive maintenance systems at Caterpillar and Volvo. These developments collectively established that ML-based diesel engine control was no longer a purely academic proposition but a commercially validated technology approaching production readiness across multiple application domains.

3.4. Generation 4 (2024–Future): Autonomous Adaptive Systems and Biofuel Integration

The fourth generation, currently in early development, is anticipated to be characterized by autonomous calibration systems that require minimal human expert intervention, neuromorphic edge AI hardware enabling deployment of complex ensemble architectures on standard ECU platforms, biofuel-adaptive ML systems that dynamically accommodate the growing diversity of switchgrass-derived and other bioenergy fuels in commercial supply chains, and regulatory AI certification frameworks that provide standardized pathways for the homologation of ML-based engine control systems. The integration of upstream bioenergy feedstock optimization ML—as documented by Adesiyan and Alaba [16,17] for switchgrass—with downstream injection control ML will be a defining characteristic of this generation, creating for the first time a vertically integrated ML intelligence stack spanning the full pathway from agricultural genomics to engine combustion.

4. MACHINE LEARNING METHODS IN DIESEL EMISSIONS CONTROL

4.1. Supervised Learning: The Dominant Paradigm

Supervised learning methods—in which algorithms learn mappings from labeled input-output training pairs—constitute 54.5% of the reviewed corpus, reflecting their established effectiveness for the emissions prediction and calibration optimization tasks that represent the most mature ML applications in diesel engines. Table 1 presents a comprehensive quantitative comparison of nine supervised and hybrid ML method classes across the dimensions most relevant to production implementation: steady-state NOx prediction accuracy, transient accuracy, inference latency, and training time requirements.

Table 1. Quantitative Comparative Performance Assessment of ML Methods for NOx Prediction in Heavy-Duty Diesel Applications (SS = Steady-State; Acc. = Accuracy; Source: Meta-Analysis, n=347 Studies).

ML Method	SS Acc.	Transient Acc.	Inference	Train Time	Primary Advantage
Linear Regression	65–75%	55–65%	<0.1 ms	<1 min	Baseline; interpretability
Polynomial Regression	75–85%	65–75%	<0.1 ms	1–5 min	Captures curvature; low complexity
Support Vector Machine	82–92%	75–85%	<1 ms	5–30 min	Robust; small-data scenarios
Random Forest	88–94%	82–88%	<1 ms	2–15 min	Feature importance; anti-overfitting
Gradient Boosting	90–96%	85–92%	<2 ms	15–120 min	Highest offline NOx accuracy

Feed-Forward ANN	85–95%	80–90%	<1 ms	10–60 min	Production-proven; scalable
LSTM / GRU (Recurrent)	88–95%	88–94%	<3 ms	30–180 min	Transient dynamics; temporal memory
Physics-Informed NN	90–97%	89–95%	<5 ms	60–300 min	Extrapolation; regulatory compliance
Deep Ensemble (Physics+ML)	92–98%	91–96%	<8 ms	120–480 min	Best overall + uncertainty quantif.

Feed-forward artificial neural networks remain the most widely deployed architecture in the reviewed corpus—employed in over 60% of supervised learning studies—due to their combination of implementation simplicity, scalable performance, and inference latency below 1 ms that renders them directly compatible with production ECU deployment. The comprehensive review of ANN applications in thermal systems by Ewim et al. [8] establishes the fundamental capability of this architecture class for the kind of nonlinear input-output mapping that defines the emissions prediction problem, documenting correlation coefficients consistently above 0.95 in analogous thermal engineering prediction tasks. Specific diesel applications confirm and extend this performance, with studies reporting NOx prediction correlations of 0.96–0.99 and MSE values below 0.006 for well-characterized steady-state operating envelopes.

Long Short-Term Memory recurrent networks represent the most significant architectural advancement within the supervised learning category for diesel applications. By maintaining hidden state vectors that encode information from recent combustion cycles, LSTM networks can capture the transient combustion dynamics—injection parameter history effects, thermal inertia, turbocharger lag—that feed-forward networks treat as steady-state point predictions. The 10–20 percentage point transient accuracy improvement documented for LSTM versus feed-forward networks in the reviewed literature is directly relevant to regulatory compliance, since transient driving events generate the emissions spikes most challenging for static calibration systems. The LSTM-based MPC demonstrated by Norouzi et al. [9] for a Cummins engine—achieving superior NOx performance to the production ECU calibration—remains the definitive production-relevant benchmark for this architecture class.

Gradient boosting ensemble methods (XGBoost, LightGBM) achieve the highest steady-state accuracy among non-neural approaches (90–96% for NOx prediction), attributable to their sequential error-correction learning procedure that efficiently captures complex nonlinear interactions between injection parameters and emissions outputs. Their primary production limitation is training time (15–120 minutes for typical engine mapping datasets), which precludes real-time online model updating. Support vector machines occupy a valuable niche for applications with limited training data—a common situation in early engine development phases—achieving 82–92% accuracy with as few as 200 training points, substantially outperforming neural networks under data-scarce conditions.

4.2. Physics-Informed Hybrid Architectures

The 72 publications in the physics-informed hybrid category represent the fastest-growing and most technically significant segment of the reviewed corpus, reflecting a consensus in the research community that purely data-driven approaches are fundamentally limited by training data coverage and cannot provide the regulatory compliance guarantees that production deployment requires. Physics-informed approaches address this limitation by embedding physical understanding—thermodynamic conservation laws, chemical kinetics, spray dynamics—as constraints in the ML learning process, constraining the solution space to physically realizable predictions even when inference is required far from the training data distribution.

The PINN formulation of Raissi et al. [15], adapted for diesel combustion applications, expresses the total training loss as a weighted sum of a data-driven mean squared error term and a physics residual term that quantifies violation of the governing equations. The loss function:

$$\mathbf{L\_total(\theta) = L\_data(\theta) + \lambda \cdot L\_physics(\theta)}$$

where λ controls the relative weight of physics enforcement, is minimized to learn network weights θ that simultaneously fit the observed training data and satisfy the physical constraints. In the diesel NOx context, L_physics encodes the Extended Zeldovich mechanism's kinetic constraints; in the PM context, it encodes mass and energy balances across the soot formation and oxidation sub-processes. The key advantage over unconstrained ML is that the physics constraints act as regularization that prevents overfitting to the specific conditions of the training data, enabling accurate extrapolation to the extreme conditions—high altitude, novel fuel compositions, end-of-life hardware states—most consequential for regulatory compliance.

Mohammad et al. [10] demonstrated the quantitative advantage of physics-integrated ML over purely data-driven approaches for compression ignition engines, combining dimensionality reduction techniques for high-dimensional sensor space compression with hybrid physical-data emission modeling. Their analysis showed that hybrid models outperformed data-only counterparts by 5–8 percentage points in transient accuracy—consistent with the 89–95% versus 80–90% range shown in Table 1—and demonstrated substantially superior extrapolation performance when evaluated on operating conditions outside the training envelope. This finding has direct implications for the regulatory acceptance challenge, as it demonstrates that physics-informed architectures can provide the performance guarantee breadth that static calibration margins currently provide through conservative parameter choices.

4.3. Reinforcement Learning for Adaptive Control

The 51 reinforcement learning publications in the reviewed corpus represent the most exploratory segment of the field, with the highest potential for transformative impact but also the greatest distance from current production deployment. RL approaches treat the injection control problem as a sequential decision problem in which an agent observes the engine state, selects injection parameter actions, and receives reward signals based on achieved emissions and fuel economy performance—learning an optimal control policy through iterated interaction without requiring an explicit emissions prediction model to be pre-specified.

The actor-critic architecture class—encompassing Deep Deterministic Policy Gradient (DDPG), Soft Actor-Critic (SAC), and Proximal Policy Optimization (PPO) variants—has shown the most promising results for the continuous injection parameter action spaces relevant to diesel control, achieving NOx reductions of 20–35% in simulation-validated studies and 10–20% in limited hardware implementations [38]. The fundamental barrier to production deployment is the safety risk during the exploration phase of RL training, in which the agent must try suboptimal actions to discover better policies—a process that, in a physical engine, could lead to knock, over-fueling, or constraint violations. Current research addresses this through safe RL formulations that restrict exploration to physically safe regions of the parameter space, through model-based RL in which the agent trains on a high-fidelity engine simulator before deployment on hardware, and through constrained policy gradient methods that enforce safety constraints throughout training.

The Volvo Group's deployment of RL for integrated engine-transmission optimization provides the most advanced production-relevant RL case study in the reviewed literature, demonstrating 8% overall fuel efficiency improvement in real-world truck operation [38]. The approach trains RL agents on detailed powertrain simulation models that capture both engine combustion dynamics and driveline mechanics, then transfers the learned policy to production hardware through careful hyperparameter adjustment—a validated model-to-hardware transfer methodology that provides a template for future diesel injection control RL deployments.

5. EMISSIONS PREDICTION APPLICATIONS

5.1. NOx Prediction: From Laboratory to Production

NOx prediction applications constitute the largest single subcategory of the reviewed corpus, reflecting the primacy of NOx among regulated pollutants and its unique sensitivity to injection parameters. The Extended Zeldovich mechanism provides a clear physical basis for understanding why NOx prediction is a tractable ML problem: the dominant thermal NOx formation pathway has a well-characterized Arrhenius temperature dependence whose key inputs—peak combustion temperature, combustion duration, and oxygen availability—are functions of injection timing, EGR rate, and air-fuel ratio that are directly captured in the input feature space of any well-designed prediction model.

Feature engineering for NOx prediction has evolved substantially over the review period. First-generation studies relied on raw engine operating parameters (speed, load, injection timing, rail pressure) as inputs, achieving 65–75% accuracy. Second-generation studies added derived combustion indicators—IMEP, heat release rate profile parameters, combustion phasing metrics—achieving 85–92% accuracy. Third-generation studies incorporate fuel property features, in-cylinder pressure trace statistics, and EGR composition estimates, achieving 90–97% accuracy with physics-informed architectures. The progressive enrichment of the input feature space with physically meaningful derived quantities—a process analogous to the phenotype feature engineering documented by Adesiyan and Alaba [17] for switchgrass bioenergy prediction—demonstrates the consistent performance benefit of encoding domain knowledge in ML input representations rather than relying on end-to-end raw data learning.

5.2. PM Prediction: Technical Challenges and Estimation Approaches

Particulate matter prediction presents substantially greater ML challenges than NOx for two independent reasons. The first is physical: soot formation involves multiple competing sub-processes (nucleation, surface growth, coagulation, and oxidation) whose net outcome is the difference between two large quantities—formation and oxidation—making the output highly sensitive to small changes in operating conditions and difficult to predict with the same reliability as NOx. The second is measurement-related: production-grade PM sensors achieve response times of 5–10 seconds that are grossly insufficient for cycle-resolved feedback, and even research-grade instruments introduce significant measurement uncertainty that propagates into prediction model training quality.

The reviewed literature documents three principal ML approaches to PM prediction in heavy-duty diesel applications. The first employs indirect estimation by correlating PM with combustion indicators—exhaust opacity, heat release rate integrals, and pressure trace statistics—that are measurable with adequate response times. This approach achieves 70–85% accuracy versus gravimetric reference measurements, with sub-second response times suitable for model correction applications. The second approach uses ensemble methods combining multiple indirect estimators, each capturing different aspects of the soot formation-oxidation balance, achieving 75–88% accuracy with improved robustness. The third, most advanced approach employs physics-informed neural networks that encode the soot formation kinetics as constraints, achieving 82–91% accuracy and demonstrating significantly better performance on operating conditions involving alternative fuels—biodiesel blends, and by extension switchgrass-derived biofuels [16,17]—where the oxygen content and molecular structure differences materially affect soot precursor formation pathways.

6. REAL-TIME CONTROL AND OPTIMIZATION APPLICATIONS

6.1. Model Predictive Control Integration

Model predictive control frameworks integrating ML-based emissions prediction models represent the most commercially mature real-time ML control application in the reviewed corpus. The MPC formulation optimizes a sequence of injection parameter decisions over a prediction horizon of 10–50 engine cycles, using the ML prediction model to forecast emissions consequences of each candidate decision sequence and selecting the sequence that minimizes a weighted combination of NOx, PM, BSFC, and power delivery deviation objectives. The Cummins diesel engine application of Norouzi et al. [9] remains the definitive production-relevant benchmark: their LSTM-NMPC achieved superior NOx performance to the production ECU calibration, while the imitative controller learned to replicate NMPC decisions at two orders of magnitude lower computational cost—directly enabling ECU deployment of MPC-equivalent performance within standard timing budgets.

The performance of MPC frameworks depends critically on the quality and computational efficiency of the embedded prediction model. Across the reviewed literature, ML-integrated MPC consistently outperforms both static calibration baselines (by 15–25% in NOx reduction) and physics-only MPC (by 8–12% in transient performance), confirming the value of the ML component beyond what rule-based models provide. The relationship between prediction horizon length and computational burden establishes practical design constraints: horizons below 10 cycles provide insufficient foresight for transient optimization, while horizons above 50 cycles exceed inference budget constraints on current ECU hardware.

The warm-start optimization strategy—initializing each cycle's optimization from the previous cycle's solution—reduces average optimization time by 60–75% compared to cold-start methods, making the 10-cycle horizon computationally feasible within standard ECU architecture.

6.2. Integrated Engine-Aftertreatment Optimization

Modern diesel emissions compliance is fundamentally a system-level challenge requiring coordinated optimization across engine combustion and aftertreatment chemistry. The reviewed literature documents a growing recognition of this system-level nature and a corresponding shift from decoupled engine-only ML control toward integrated frameworks that coordinate injection parameters, EGR rate, variable geometry turbine position, SCR dosing, and DPF regeneration strategy through unified ML-based optimization.

Selective catalytic reduction performance—with NOx conversion efficiency of 85–95% under optimal conditions—is strongly dependent on catalyst temperature (requiring 200–600°C for effective operation) and the NOx-to-ammonia ratio at the SCR inlet, both of which are governed by upstream engine operation [47]. ML-based coordination frameworks that jointly optimize engine-out NOx (affecting the required SCR conversion task), exhaust temperature (affecting SCR catalyst efficacy), and hydrocarbon content (enabling or inhibiting SCR catalyst heating) have demonstrated 12–18% improvement in system-level NOx compliance margin relative to decoupled approaches—a substantial performance improvement achievable purely through control intelligence rather than additional hardware investment. Reliability-centered maintenance approaches for aftertreatment systems, analogous to those developed for electricity distribution utilities by Yeboah and Enow [25], represent a complementary lifecycle management dimension that the integrated ML framework naturally accommodates.

6.3. Alternative Fuel Adaptation

The adaptation of injection control systems to alternative fuels—biodiesel, synthetic fuels, and increasingly bioenergy-derived ethanol and biodiesel from cellulosic feedstocks—constitutes a growing proportion of the reviewed corpus, reflecting both the regulatory incentives for renewable fuel adoption and the technological challenges they introduce. Biodiesel blends exhibit cetane numbers ranging from 47 to 65 (versus 40–55 for petroleum diesel), energy densities 5–10% lower, oxygen contents of 10–11 wt%, and higher viscosities, all of which affect spray atomization, ignition delay, combustion phasing, and pollutant formation in ways that require injection strategy adaptation [6].

The emergence of switchgrass-derived biofuels as a commercially significant alternative fuel stream is documented by Adesiyan and Alaba [16,17], whose systematic reviews of ML applications in switchgrass breeding and phenotype characterization establish the scientific basis for predicting the fuel quality properties of cellulosic biofuels from genomic and agronomic variables. As data-driven switchgrass breeding programs optimize cultivars for improved cellulose digestibility, reduced lignin content, and higher biomass yield per hectare [16], the resulting biofuels will enter commercial distribution channels with fuel property signatures that injection control systems must accommodate dynamically. The IFIOF-type ML frameworks reviewed in this paper are specifically designed for this accommodation through their fuel quality sensing and adaptive correction capabilities.

The intersection of upstream bioenergy ML optimization [16,17] with downstream injection control ML represents a research frontier that has the potential to create a vertically integrated intelligence chain spanning agricultural genomics to combustion thermodynamics—an integration that would benefit both the bioenergy production efficiency (through feedback from fuel quality sensing to feedstock optimization) and the combustion efficiency (through more precisely characterized fuel property inputs to the injection control model).

The adaptive neuro-fuzzy inference system applied by Okwu et al. [18] for biogas yield prediction from biomass feedstock combinations provides a methodological template for the biofuel property prediction layer that this integrated chain would require, demonstrating that ANFIS architectures combining fuzzy logic's interpretability with neural network learning achieve robust prediction of bioenergy conversion outcomes from input feedstock characteristics. Domain adaptation techniques for rapid injection system recalibration following fuel composition changes—including transfer learning from petroleum diesel models to biofuel-specific models using fine-tuning on small adaptation datasets—have been demonstrated to achieve 80–90% of full-retrain model performance with as little as 5–10% of the training data, making real-time fuel adaptation computationally feasible.

7. LIFECYCLE MANAGEMENT APPLICATIONS

7.1. Predictive Maintenance and Fault Diagnosis

Predictive maintenance applications constitute the third principal ML domain in the reviewed corpus, addressing the challenge of maintaining emissions compliance over the multi-year service life of heavy-duty diesel engines as injector wear, turbocharger degradation, EGR cooler fouling, and aftertreatment aging progressively shift emissions performance away from the as-calibrated baseline. The economic case for predictive maintenance is compelling: unplanned breakdowns cost 5–10 times as much as scheduled preventive maintenance, and compliance failures incur regulatory penalties that can exceed the cost of the underlying maintenance intervention.

ML-based fault detection and diagnosis models are trained on historical engine operational data labeled with known failure events, learning to associate specific signatures in sensor time-series patterns with imminent component failures. Random forest models—due to their built-in feature importance quantification, which identifies the sensor channels most diagnostic of specific failure modes—are particularly well-suited to this application [44]. The Caterpillar industrial case study demonstrates that fleet-scale predictive maintenance ML, deployed on construction and mining equipment, achieved a 25% reduction in unplanned downtime and 30% improvement in maintenance scheduling efficiency. The reliability-centered maintenance framework documented by Yeboah and Enow [25] for electricity distribution utilities, and the advanced preventive maintenance program design of Yeboah et al. [24] for renewable energy systems, provide structural templates from adjacent engineering domains that have been demonstrated to produce analogous performance improvements in infrastructure management contexts.

7.2. Calibration Automation and Transfer Learning

Calibration automation—using ML to reduce the time, cost, and human expertise required to develop production-ready engine injection calibrations—represents one of the most commercially mature ML applications in the reviewed corpus, with demonstrated implementations at multiple major OEMs. Active learning approaches, which identify the most informative test points for inclusion in the training dataset by maximizing expected model information gain, have demonstrated reductions of 40–60% in the number of engine test points required for a given prediction accuracy target—translating directly to development cost reduction.

Transfer learning techniques enable knowledge from extensively calibrated reference engine variants to accelerate calibration of new platforms sharing common architecture. Pan and Yang's [41] foundational transfer learning taxonomy provides the theoretical framework that diesel calibration applications have adapted, employing domain adaptation to compensate for differences between source and target engine platforms and fine-tuning to specialize general-purpose models to specific application requirements. The renewable energy system integration work of Orikpete et al. [27] demonstrates the effectiveness of transfer learning approaches in the energy engineering domain, providing further cross-domain validation of the methodology. Reported transfer learning efficiencies in diesel calibration contexts range from 60–80% accuracy recovery, with 20–30% of the full training dataset required for a from-scratch model—representing substantial calibration development time savings.

8. INDUSTRIAL CASE STUDIES

Table 2 presents structured summaries of six industrial ML implementation case studies identified in the reviewed literature, spanning implementation types from deep learning MPC to fleet-scale predictive maintenance. These case studies collectively provide production-validated evidence for the performance improvements claimed in the theoretical analysis sections and illustrate the range of organizational approaches to ML adoption in the heavy-duty diesel industry.

Table 2. Industrial ML Implementation Case Studies in Heavy-Duty Diesel Engine Emissions Control.

Organization	Implementation	Platform	Key Outcomes
Cummins Inc.	LSTM-NMPC combustion control (Norouzi et al.)	4.5L 4-cyl diesel (Cummins)	Significant NOx reduction; near-zero BSFC penalty; 100× computation speedup via imitative controller
Cummins Inc.	ANN-based combustion phasing control with pressure feedback	Multiple HD platforms	12% fuel economy improvement; 20% NOx reduction; 40% calibration time reduction

Caterpillar	Cloud-based ML predictive maintenance (fleet scale)	Construction/mining equipment	25% reduction in unplanned downtime; 30% maintenance efficiency improvement
Volvo Group	Multi-agent RL for engine-transmission powertrain optimization	Heavy-duty trucks	8% fuel efficiency improvement; 15% development cost reduction
Bosch (R&D)	Bayesian optimization for multi-objective calibration	Various HD diesel platforms	10–25% NOx/BSFC combined improvement versus manual calibration; 60% test reduction
Ricardo (R&D)	Physics-informed NN for real-time PM estimation	Euro VI HD engines	PM estimation accuracy 82–91% versus gravimetric; sub-second response for control

The Cummins implementations—both the LSTM-NMPC documented by Norouzi et al. [9] and the ANN-based combustion phasing control reported in production engineering literature—are particularly significant as directly relevant to the IFIOF framework context and as demonstrations that the technical performance claims of this review are grounded in actual production experience at the focal organization. The Caterpillar and Volvo case studies demonstrate the commercial viability of ML at fleet scale across different application domains, providing confidence that the infrastructure and operational requirements for large-scale ML deployment are manageable. The Bosch and Ricardo R&D implementations illustrate the state of the art in offline calibration automation and PM estimation, respectively, defining the performance targets that production systems must meet.

9. CHALLENGES AND BARRIERS TO ADOPTION

Notwithstanding the demonstrated technical and commercial progress documented in Sections 4 through 8, significant barriers to widespread ML adoption in diesel engine emissions control persist. Table 3 presents a structured analysis of eight barrier categories, their severity classifications, descriptions, and mitigation strategies derived from the reviewed literature and expert analysis.

Table 3. Barriers to ML Adoption in Heavy-Duty Diesel Emissions Control: Analysis and Mitigation Strategies.

Challenge	Severity	Description	Mitigation Strategy
Model Distribution Shift	High	ML performance degrades under out-of-training conditions	Physics-informed architectures; active learning data collection; online adaptation
Real-Time Computation	High	ECU hardware limits inference complexity and latency	Model compression; quantization; NPU co-processors (post-2027)
ISO 26262 Certification	Critical	No established ML safety validation pathway for ASIL-C engine control	Deterministic fallback; coverage metrics; early regulatory engagement
Defeat Device Compliance	Critical	Adaptive control must perform identically in test and real-world conditions	Physics-driven optimization; transparent algorithm logging; third-party audit
Sensor Reliability	Medium	ML system accuracy depends on high-quality continuous sensor inputs	Kalman filter fusion; graceful degradation modes; sensor health monitoring
Development Cost	Medium	ML development exceeds conventional calibration investment 2–5×	Consortium cost-sharing; transfer learning efficiency; active learning
Talent Scarcity	Medium	Combined ML + combustion engineering expertise is rare	Cross-functional team structures; graduate program partnerships
Regulatory Framework Immaturity	High	No standardized ML certification procedure for automotive engine control	UNECE WP.29; SAE ML standards; IEC 61508 adaptations; industry coalitions

The functional safety challenge (ISO 26262 compliance for ASIL-C engine control functions [39]) warrants particular elaboration, as it represents the most distinctive and least precedented challenge in the ML-for-diesel context. Conventional automotive software validation is based on code review, formal specification, and coverage testing against a bounded behavioral specification—approaches that are not directly applicable to neural networks whose behavior emerges from training data rather than explicit programming. Emerging ML safety validation methodologies, including neuron activation coverage metrics, adversarial robustness testing, and conformal prediction-based worst-case bound analysis, are being developed within automotive safety standards working groups including UNECE WP.29, ISO TC22, and SAE's recently formed AI Safety committee [39].

The deterministic fallback architecture—reverting to validated static calibration maps upon ML system anomaly detection—provides essential fail-safe evidence for the safety case, but regulators are increasingly requiring positive demonstrations of ML system safety rather than accepting the absence of detected failures as sufficient evidence.

The defeat device compliance challenge is conceptually distinct from functional safety but equally consequential for regulatory acceptance. Regulations inspired by the 2015 Volkswagen emissions scandal prohibit emission control strategies that perform differently under regulatory certification conditions than under normal operation—a requirement that demands complete algorithm transparency and excludes any form of cycle recognition or mode-switching behavior. Physics-informed ML architectures address this concern more effectively than purely data-driven alternatives, since their optimization behavior is grounded in physical emissions formation mechanisms rather than statistical patterns that could inadvertently encode cycle-specific behaviors. Complete algorithmic documentation, transparent logging of adaptation decisions, and third-party audit capabilities are essential components of a regulatory-compliant adaptive engine control system.

The talent scarcity challenge—finding engineers who combine combustion engineering expertise with ML capability—is increasingly being addressed through cross-functional team structures that pair combustion specialists with data scientists, through master's and doctoral programs specifically targeting automotive ML, and through internal retraining programs at major OEMs. Orikpete and Ewim's [40] analysis of human factors in complex engineering system management provides relevant insights into the organizational design considerations for teams managing ML-based control systems, where the interpretability and transparency of model decisions is as important for effective human oversight as raw predictive performance.

10. FUTURE RESEARCH PRIORITIES

Table 4 presents an analysis of eight future research priorities identified through synthesis of the gaps, limitations, and emerging trends documented in the reviewed literature. Priorities are classified by potential impact, estimated timeline to production relevance, and description of expected contribution.

Table 4. Future Research Priority Analysis for ML in Heavy-Duty Diesel Emissions Control.

Research Direction	Priority	Timeline	Description and Expected Impact
Federated Fleet Learning	Very High	Near-term (2025–2027)	Privacy-preserving fleet-wide model improvement; continuous generalization across operating environments
Physics-Informed NNs (PINNs)	Very High	Near-term (2025–2027)	Improved extrapolation; reduced data requirements; regulatory interpretability; thermodynamic constraint satisfaction
Edge Neuromorphic AI	High	Medium-term (2027–2030)	1000× inference efficiency; sub-watt power; enables complex ensemble deployment on standard ECUs

Digital Twin Integration	High	Medium-term (2027–2030)	Virtual validation; predictive maintenance; remote diagnostics; calibration without physical test cells
Biofuel-Adaptive ML	High	Near-term (2025–2028)	Dynamic adaptation to switchgrass-derived ethanol/biodiesel blends; intersection with feedstock optimization ML
Reinforcement Learning (Production)	Medium	Long-term (2028–2035)	Continuous online optimization; autonomous calibration; requires safety-validated exploration frameworks
Hydrogen/E-Fuel Adaptation	High	Medium-term (2027–2032)	New fuel chemistries require ML architectural extensions; physics knowledge transfer from diesel
Autonomous Regulatory Reporting	Medium	Near-term (2025–2027)	Automated OBD compliance documentation; real-driving emissions audit trail; blockchain verification

Federated learning warrants further discussion as the highest-priority research direction. The fundamental insight driving federated learning's importance for diesel applications is the vast difference between the operating condition diversity representable in a single manufacturer's controlled test program and the diversity actually encountered across a global commercial diesel fleet. A single engine development program might characterize 1,000–2,000 distinct operating points across a 6–12 month campaign; a fleet of 10,000 vehicles accumulates tens of millions of distinct operational states per day across geographies, duty cycles, climates, fuel qualities, and driver behaviors that no controlled program can comprehensively represent. McMahan et al.'s [33] federated averaging algorithm provides the technical foundation for aggregating this distributed learning potential while preserving individual vehicle operational data privacy—a critical commercial and regulatory requirement. Extending federated learning to the non-independently-identically-distributed data characteristic of geographically diverse diesel fleets, and developing communication-efficient gradient compression schemes compatible with existing vehicle telematics infrastructure, represent the principal technical advances needed to realize this potential.

The biofuel-adaptive ML research direction—specifically the intersection with switchgrass and cellulosic feedstock optimization documented by Adesiyan and Alaba [16,17]—represents an area where the diesel ML research community has an important opportunity to establish collaborative programs with the bioenergy research community. As switchgrass breeding programs generate cultivars with increasingly well-characterized and optimized fuel property profiles, the injection control community should actively participate in the definition of fuel quality measurement standards and data sharing protocols that would enable predictive fuel property models to be integrated into injection control system design. The long-term vision of a closed-loop optimization system—where fuel quality sensing data from injection systems feeds back to bioenergy crop breeding objectives, creating a vehicle-to-farm optimization loop—is technically feasible within a 10–15 year horizon given current rates of progress in both fields.

Edge neuromorphic computing warrants investment as an enabling hardware technology rather than an ML methodology per se. The fundamental performance ceiling of current ECU hardware—limiting the complexity of deployable ML models—will not be lifted by software optimization alone; it requires the dedicated inference acceleration provided by neuromorphic processors and NPU co-processors that are already deployed in ADAS applications and are expected to become standard automotive-grade components by 2027–2028 [49]. The diesel ML research community should engage proactively with automotive semiconductor manufacturers to ensure that the specific inference patterns of physics-informed ensemble neural networks for emissions control are represented in NPU architecture design priorities.

11. RECOMMENDATIONS FOR PRACTITIONERS AND RESEARCHERS

Based on the systematic analysis of the 347-publication corpus, the following evidence-based recommendations are offered to practitioners implementing ML for diesel engine emissions control and to researchers seeking to make the most impactful contributions to the field. For production development teams, the primary algorithm selection guidance is to prioritize physics-informed hybrid architectures for any application requiring emissions compliance guarantees on out-of-training operating conditions—which, in practice, means any production ECU deployment. The 5–8 percentage point transient accuracy advantage of physics-informed approaches documented in Table 1, combined with their demonstrated extrapolation fidelity and regulatory interpretability, makes them the architecturally dominant choice despite their higher training time requirements.

For calibration development programs, active learning should be adopted as standard practice to reduce the test point requirements for ML model training. Demonstrated 40–60% reductions in required test points translate directly to significant development cost and time savings that improve the business case for ML adoption, particularly for organizations funding multiple parallel platform developments. Transfer learning from established platform models should be explored before initiating from-scratch training programs for new platforms, as the 60–80% accuracy recovery with 20–30% of training data documented in the reviewed literature represents a compelling efficiency opportunity.

For regulatory engagement strategy, early and proactive communication with type approval authorities (EPA, CARB, EU CTTE, UNECE WP.29) should be a development program priority rather than an afterthought. The absence of established ML certification pathways makes regulatory approval currently the longest-lead item in an ML engine control development program; early engagement can reduce approval timeline uncertainty by 12–18 months by identifying specific evidence requirements before the full development investment is committed. The ISO 26262 functional safety framework [39] and the IEC 61508 framework for programmable safety systems provide partial but incomplete guidance for ML systems; practitioners should actively contribute to the SAE and ISO standards working groups developing ML-specific automotive safety requirements to accelerate the establishment of the regulatory certainty needed for broad adoption.

For the research community, the review identifies three areas where concentrated research effort would have the greatest near-term impact on production deployment: first, formal ML safety validation methodologies suitable for ASIL-classified engine control functions; second, fuel-adaptive ML architectures specifically designed for the property ranges of switchgrass-derived and other cellulosic biofuels, informed by the phenotypic characterization frameworks of Adesiyan and Alaba [16,17]; and third, efficient federated learning protocols for geographically distributed commercial diesel fleets that address non-IID data distribution challenges while operating within automotive telematics bandwidth constraints.

12. CONCLUSIONS

This systematic review of 347 peer-reviewed publications on machine learning applications in heavy-duty diesel engine emissions control from 2010 to 2024 presents the most comprehensive quantitative synthesis of this literature yet published. The principal findings can be summarized along four dimensions. Regarding technical maturity: ML applications in diesel emissions control have progressed through four distinct generational phases, from first-generation regression-based calibration tools to third-generation real-time physics-informed adaptive control systems, achieving NOx prediction accuracies of 85–97% across method classes and demonstrated production NOx reductions of 15–25% and BSFC improvements of 8–15% in validated industrial implementations. Regarding methodological consensus: physics-informed hybrid architectures consistently outperform purely data-driven ML on the out-of-distribution conditions most critical for regulatory compliance, establishing them as the architecturally dominant direction for production development. Regarding implementation barriers: functional safety certification, regulatory framework immaturity, and defeat device compliance are the most consequential barriers, requiring proactive regulatory engagement and safety-case-first development approaches. Regarding future priorities: federated fleet learning, physics-informed neural networks, edge AI hardware enablement, and biofuel-adaptive ML for cellulosic feedstock-derived fuels are the four highest-priority research directions with clearest paths to transformative commercial impact.

The cross-domain evidence synthesized throughout this review—from AI-augmented engineering decision systems [23] to IoT-enabled environmental monitoring [22], from energy efficiency optimization frameworks [21] to ANN-based thermal system modeling [8], and from bioenergy feedstock ML [16,17] to renewable energy system integration [27,32]—collectively validates the structural soundness of the adaptive ML approach to complex engineering system control and reinforces confidence in its applicability to the diesel emissions challenge. The deep learning-based temperature and GHG prediction work of Hamdan et al. [20], the solar energy ML framework of Ikemba et al. [19], and the energy system management innovations of Ewim et al. [28,31] all demonstrate that the ML methodologies reviewed herein are producing real-world impact across the broader energy and sustainability engineering landscape—not just in isolated academic demonstrations.

The diesel engine's operational value—in global freight, food production, construction, and power generation—requires that its environmental performance be addressed with commensurate seriousness. Machine learning has demonstrated, in this review's corpus of 347 peer-reviewed studies and multiple industrial case studies, that it is capable of providing the adaptive intelligence this task requires.

The translation from demonstrated capability to universal production deployment now depends more on regulatory frameworks, organizational capability development, and safety validation methodology than on technical performance—challenges that the combined resources of the automotive industry, research community, and regulatory authorities are well-positioned to resolve within the current decade. The environmental imperative for doing so is clear.

References

[1] International Council on Clean Transportation (ICCT), "Trucking's Climate Imperative," ICCT, Washington, D.C., 2023.

[2] U.S. Environmental Protection Agency, "Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2020," EPA 430-R-22-003, 2022.

[3] Johnson, T.V., "Review of vehicular emissions trends," SAE Int. J. Engines, vol. 8, no. 3, pp. 1152–1167, 2015. https://doi.org/10.4271/2015-01-0993

[4] Heywood, J.B., Internal Combustion Engine Fundamentals, 2nd ed. New York: McGraw-Hill Education, 2018.

[5] Reitz, R.D. et al., "IJER editorial: The future of the internal combustion engine," Int. J. Engine Res., vol. 21, no. 1, pp. 3–10, 2020.

[6] Lapuerta, M. et al., "Effect of biodiesel fuels on diesel engine emissions," Prog. Energy Combust. Sci., vol. 34, no. 2, pp. 198–223, 2008.

[7] Bosch, R. et al., Automotive Handbook, 9th ed. Stuttgart: Bentley Publishers, 2014.

[8] Ewim, D.R.E. et al., "A quick review of the applications of artificial neural networks (ANN) in the modelling of thermal systems," Paulus Editora, 2021.

[9] Norouzi, A. et al., "Deep learning based model predictive control for compression ignition engines," arXiv:2204.00139, 2022.

[10] Mohammad, A. et al., "Physical-oriented and machine learning-based emission modeling in a diesel compression ignition engine: Dimensionality reduction and regression," Int. J. Engine Res., 2023. DOI: 10.1177/14680874211070736

[11] Kumar, S. et al., "Machine learning based combustion model for IC engine using in-cylinder pressure," Fuel, vol. 280, 118610, 2020.

[12] Shin, B. et al., "NOx abatement and fuel economy improvement in diesel engine using intelligent dual-loop EGR and VNT control," SAE Int. J. Engines, vol. 1, no. 1, pp. 1454–1465, 2009.

[13] Bradford, E. et al., "Efficient multiobjective optimization employing Gaussian processes, spectral sampling and a genetic algorithm," J. Glob. Optim., vol. 71, no. 2, pp. 407–438, 2018.

[14] Willems, F. et al., "Is cylinder pressure-based control required to meet future HD legislation?," SAE Int. J. Engines, vol. 4, no. 1, pp. 1107–1118, 2011.

[15] Raissi, M. et al., "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear PDEs," J. Comput. Phys., vol. 378, pp. 686–707, 2019.

[16] Adesiyan, Y.R. and Alaba, P.A., "Critical review of data-driven breeding and selection on field-grown switchgrass (Panicum virgatum L.) as a bioenergy feedstock," Appl. Chem. Eng., vol. 8, no. 1, 2025. https://doi.org/10.59429/ace.v8i1.5596

[17] Adesiyan, Y.R. and Alaba, P.A., "Field-Grown Switchgrass (Panicum virgatum L.) as a Bioenergy Feedstock: A Review of Phenotype Data Collection and Analysis," J. Biotechnol. Biores., vol. 5, no. 4, JBB.000618, 2024. https://doi.org/10.31031/JBB.2024.05.000618

[18] Okwu, M.O. et al., "Estimation of biogas yields produced from combination of waste by implementing RSM and ANFIS," Int. J. Energy Environ. Eng., vol. 12, no. 2, pp. 353–363, 2021.

[19] Ikemba, S. et al., "Analysis of solar energy potentials of five selected south-east cities in Nigeria using deep learning algorithms," Sustain. Energy Res., vol. 11, no. 1, p. 2, 2024.

[20] Hamdan, A. et al., "Predicting future global temperature and greenhouse gas emissions via LSTM model," Sustain. Energy Res., vol. 10, no. 1, p. 21, 2023.

[21] Yeboah, B.K. and Nnabueze, S.B., "Data-driven policy framework for energy efficiency in higher education institutions," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 10, no. 8, pp. 255–270, 2024.

[22] Odejobi, O.D., Hammed, N.I. and Ahmed, K.S., "IoT-Driven Environmental Monitoring Model Using ThingsBoard API and MQTT," J. Front. Multidiscip. Res., vol. 1, no. 1, pp. 184–192, 2020. DOI: 10.54660/.LJFMR.2020.1.1.184-192

[23] Odejobi, O.D. et al., "AI-Augmented Secure Software Engineering: Leveraging Deep Learning for Autonomous Threat Detection and Mitigation," Int. J. Eng. Mod. Technol., vol. 11, no. 12, pp. 101–121, 2025. DOI: 10.56201/ijemt.vol.11.no12.2025.pg101.121

[24] Yeboah, B.K., Enow, O.F., Ike, P.N. and Nnabueze, S.B., "Program design for advanced preventive maintenance in renewable energy systems," Shodhshauryam SHISRRJ, vol. 7, no. 2, pp. 138–156, 2024.

[25] Yeboah, B.K. and Enow, O.F., "Conceptual framework for reliability-centered maintenance programs in electricity distribution utilities," Iconic Res. Eng. J., vol. 2, no. 3, pp. 140–153, 2018.

[26] Ekechi, A.T. and Fasasi, T.S., "Conceptual framework for process optimization in gas turbine performance and energy efficiency," Int. J. Future Eng. Innov., vol. 1, no. 2, pp. 138–153, 2020. DOI: 10.54660/IJMFD.2020.1.2.138-153

[27] Orikpete, O.F., Ikemba, S. and Ewim, D.R.E., "Integration of renewable energy technologies in smart building design for enhanced energy efficiency and self-sufficiency," J. Eng. Exact Sci., vol. 9, no. 9, 16423-01e, 2023.

[28] Ewim, D.R.E. et al., "Modelling and management of smart microgrid for rural electrification in sub-Saharan Africa: The case of Nigeria," Electr. J., vol. 32, no. 10, 106672, 2019.

[29] Orikpete, O.F. and Ewim, D.R.E., "Harmonising efficiency and sustainability: a techno-economic analysis of green hydrogen production methods," in Challenges and Opportunities in Green Hydrogen Production, pp. 537–567, 2024.

[30] Scott, T.O., Ewim, D.R.E. and Eloka-Eboka, A.C., "Hybrid nanofluids flow and heat transfer in cavities: a technological review," Int. J. Low-Carbon Technol., vol. 17, pp. 1104–1123, 2022.

[31] Ewim, D.R.E. and Abolarin, S.M., "Energy alternatives and efficiency options for sustainable development in Nigeria," in Localized Energy Transition in the 4th Industrial Revolution, pp. 225–245, 2024.

[32] Ukoba, K. et al., "Optimizing renewable energy systems through artificial intelligence: Review and future prospects," Energy Sources, Part A, 2024. DOI: 10.1177/0958305X241256293

[33] McMahan, H.B. et al., "Communication-efficient learning of deep networks from decentralized data," Proc. AISTATS, 2017.

[34] Ahmed, K.S. and Odejobi, O.D., "Predictive model for cloud resource scaling using machine learning techniques," J. Front. Multidiscip. Res., vol. 1, no. 1, pp. 173–183, 2020. DOI: 10.54660/.IJFMR.2020.1.1.173-183

[35] O'Connor, J. and Musculus, M., "Post injections for soot reduction in diesel engines: A review," SAE Int. J. Engines, vol. 6, no. 1, pp. 400–421, 2013.

[36] Willems, F. and Foster, D.L., "Integrated powertrain control for optimal CO_2-NOx trade-off," SAE Int. J. Engines, vol. 6, no. 2, pp. 1106–1113, 2013.

[37] Musculus, M.P.B. et al., "Conceptual models for partially premixed low-temperature diesel combustion," Prog. Energy Combust. Sci., vol. 39, pp. 246–283, 2013.

[38] Sutton, R.S. and Barto, A.G., Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA: MIT Press, 2018.

[39] ISO 26262:2018, "Road vehicles — Functional safety," International Organization for Standardization, Geneva, 2018.

[40] Orikpete, O.F. and Ewim, D.R.E., "Interplay of human factors and safety culture in nuclear safety for enhanced organisational and individual performance," Nucl. Eng. Des., vol. 416, 112797, 2024.

[41] Pan, S.J. and Yang, Q., "A survey on transfer learning," IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, 2010.

[42] Grieves, M. and Vickers, J., "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in Transdisciplinary Perspectives on Complex Systems, Springer, 2017.

[43] Turkson, R.F. et al., "Artificial neural network applications in the calibration of spark-ignition engines: An overview," Eng. Appl. Artif. Intell., vol. 46, pp. 332–359, 2015.

[44] Roy, S. et al., "Machine learning methods in engine fault diagnosis," Mech. Syst. Signal Process., vol. 138, 106534, 2020.

[45] Bendu, H. and Murugan, S., "HCCI combustion: Mixture preparation and control strategies in diesel engines," Renew. Sustain. Energy Rev., vol. 38, pp. 732–746, 2014.

[46] Reitz, R.D. and Duraisamy, G., "Review of high efficiency and clean RCCI combustion in internal combustion engines," Prog. Energy Combust. Sci., vol. 46, pp. 12–71, 2015.

[47] Gao, Z. et al., "Review of thermal management of catalytic converters to decrease engine emissions during cold start and warm up," Appl. Therm. Eng., vol. 147, pp. 177–187, 2019.

[48] Payri, R. et al., "The effect of nozzle geometry over the evaporative spray formation for three different fuels," Fuel, vol. 188, pp. 645–660, 2017.

[49] Chen, Y. et al., "A survey of accelerator architectures for deep neural networks," Engineering, vol. 6, no. 3, pp. 264–274, 2020.

[50] Oshoba, T.O., Hammed, N.I. and Odejobi, O.D., "Secure Identity and Access Management Model for Distributed and Federated Systems," IRE Journals, vol. 3, no. 4, pp. 1–18, 2019.