Scientific Reports

https://doi.org/10.1038/s41598-026-40929-1

Article in Press

Response surface and TQM-ML analysis of a PCCI engine fueled with PO and microalgae biodiesel

Received: 16 December 2025

Accepted: 17 February 2026

Published online: 23 February 2026

Cite this article as: Al Awadh M. & Michael G.K.O. Response surface and TQM-ML analysis of a PCCI engine fueled with PO and microalgae biodiesel. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-40929-1

Mohammed Al Awadh & Goh Kah Ong Michael

Response Surface and TQM-ML Analysis of a PCCI Engine Fueled with PO and Microalgae Biodiesel

Mohammed Al Awadh [1,2], Goh Kah Ong Michael [3]*

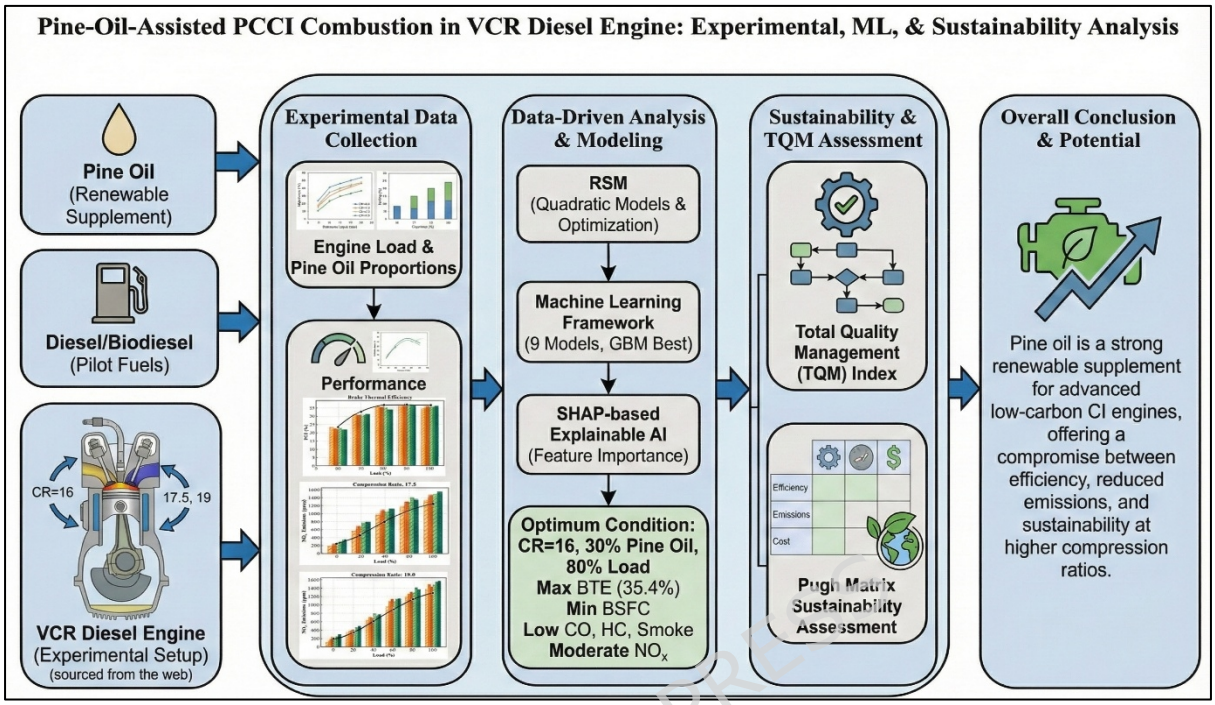
[1]Department of Industrial Engineering, College of Engineering, King Khalid University, P.O. Box 394, Abha 61421, Saudi Arabia;
[2]Center for Engineering and Technology Innovations, King Khalid University, Abha 61421, Saudi Arabia. (mohalawadh@kku.edu.sa)
[3]Center for Image and Vision Computing, COE for Artificial Intelligence, Faculty of Information Science & Technology, Multimedia University, Jln Ayer Keroh Lama, Melaka, 75450 Malaysia; michael.goh@mmu.edu.my

**Corresponding: michael.goh@mmu.edu.my*

Abstract

This work presents a comprehensive experimental and data-driven study on the feasibility of pine-oil-aided premixed charge compression ignition (PCCI) combustion under low-temperature combustion (LTC) conditions in a variable compression ratio (VCR) diesel engine using conventional diesel and biodiesel blends to operate as pilot fuel. Different amounts of Pine Oil (PO) were used at 16, 17.5, and 19 compression ratios with different engine loads during tests. Performance and emissions results such as BTE, BSFC, CO, HC, NOx, and smoke opacity were examined. RSM generated statistically significant quadratic models and was used for simultaneous multi-objective optimisation. The optimal operating condition is CR = 19 with 30% PO with 80% load. This yielded a peak BTE of 35.4%, a minimum BSFC of 0.25 kg/kWh, a CO level of 0.022%, an HC level of 31 ppm and a smoke opacity of 21 HSU. NOx: an increase (1120 ppm) was also observed. In the present work, nine regression models were employed in a framework for machine learning. Among various models, the Gradient Boosting Machine had the highest prediction accuracy ($R^2 > 0.95$). SHAP-based explainable AI revealed that engine load, compression ratio, and fuel properties were the most influential on how combustion behaved. The TQM and sustainability assessment based on the Pugh matrix indicated that the use of PO to enable operating PCCI at higher compression ratios offers the best compromise of efficiency with low emissions and sustainability between the different options. These combined outcomes indicate that PO has significant potential as a renewable fuel for advanced low-carbon compression ignition engines.

Keywords: PCCI combustion; PO; Variable compression ratio engine; Response Surface Methodology; Machine learning; SHAP analysis; Total Quality Management; Pugh matrix; Sustainability assessment; Emission characteristics.



1. Introduction

The move to stricter emission rules and the departure from fossil fuels have increased demand for renewables. Microalgae biodiesel blends have lower carbon footprints but show high viscosity and cetane number problems, which lead to low atomisation and lack of feasibility for PCCI for low-temperature combustion strategies [1]. Pine oil (PO), which is a low-viscosity and low-flammability terpene micro-emulsion fuel, has been proposed as a promising co-solvent to overcome these limitations. The availability of PO, which helps in controlling ignition delay and fuel–air homogeneity, provides a means of safe PCCI combustion with reduced NO_x–soot trade-off [2]. The ICE sector is going through a considerable transformation as tightening global emission regulations increase the shift to cleaner powertrains. While conventional compression ignition (CI) engines achieve higher thermal efficiency, they intrinsically face the NO_x–soot trade-off. To overcome this restriction, low-temperature combustion (LTC) strategies such as premixed charge compression ignition (PCCI) have received significant interest [3]. Here, early fuel injection and high premixing characterise the PCCI (Split Injection), to decouple soot and NOx formation and minimise rich zones and high-temperature regions known for soot and thermal NOx formation [4]. Despite making it such appealing, the PCCI operation is restricted to a very small operating

window by low-load combustion stability issues and high-load excessive pressure rise rates.

Recent studies have shown that the window of PCCI combustion can be widened through advanced control strategies such as high injection pressure and variable compression ratio (VCR) operation for improving fuel atomisation, mixing, and thus, combustion phasing [5]. On the other hand, the positive effect of chemical reactivity enhancement by using different fuels such as hydrogen and oxygenated biofuels have also been investigated for increasing combustion efficiency and emission performance [6]. Both NO_x and soot emissions can be significantly decreased using PCCI; however, unburned hydrocarbon (HC) emission is a persistent concern, especially at low temperature and high premixed status [7].

Biofuels containing oxygen in their molecular structure can facilitate soot oxidation under partially premixed conditions (PCCI), potentially improving PCCI engine performance. Molecular oxygen supplied by biodiesel and higher alcohols is conducive to the oxidation of soot precursors in locally rich regions where complete homogeneity is impossible [8]. But if more fuel oxygen content is desired, this has to be optimised since too much oxygenation can worsen the NO_x–soot trade-off. This has led to a recent trend in developing blended fuels that should be ideally suited to achieve the required reactivity stratification for stable PCCI operation. Especially biodiesel–alcohol blends, which have high reactivity of biodiesel and low viscosity and high volatility of alcohols, enhance atomisation, evaporation, and mixing while alleviating pressure rise rates over broader load ranges [9].

The chemical structure of oxygenated fuels is important for PCCI fuel combustion characterisation. The increased ignition delay can provide additional time for fuel–air mixing as a result of low cetane number oxygenates and reduces the propensity to form fuel-rich regions before combustion [10]. Nonetheless, these oxygenated blends have a lower energy density, making them responsible for higher BSFC [11]. As a result, recent optimisation investigations have targeted finding maximum reduction benefit ratios while keeping thermal performance equivalent to conventional diesel fuelling [12]. Pine oil (PO), a terpene-rich biofuel obtained from pine resin or as a by-product from the pulp industry, has received growing attention as a candidate of PCCI for its favourable physicochemical properties. PO has much lower viscosity (1.3 cSt) and a lower boiling point than either normal diesel or fatty acid methyl esters. Such characteristics facilitate better atomisation, and faster vaporisation to promote better fuel–air homogeneity under PCCI conditions, which helps minimise soot [2].

It is mainly composed of terpene-derived compounds such as α-terpineol and α-pinene, which provide a low cetane number (10–15). The consequent low cetane number will delay the ignition so that the premixing time will be longer, and the diffusion-controlled combustion and soot formation will be suppressed in PCCI operation (Lapuerta et al., 2023). (Zhu et al., 2023). Mixing terpene with PO has been reported to achieve virtually identical brake thermal efficiencies compared to diesel combustion and greater reductions in soot emissions and CO emissions due to the presence of fuel-bound oxygen in the terpene molecules (CHIVU 2025; Neupane 2022), for such engine configurations. It is mainly composed of terpene-derived compounds such as α-terpineol and α-pinene, which provide a low cetane number (10–15). This same property, however, can cause low-load combustion instability when using PO alone. It was recently shown that using PO to devise dual-fuel strategies in which a high-reactivity pilot fuel (diesel or biodiesel) appropriately manages ignition timing is more effective [13]. Mixing terpene with PO has been reported to achieve virtually identical brake thermal efficiencies compared to diesel combustion and greater reductions in soot emissions and [14] CO emissions due to the presence of fuel-bound oxygen in the terpene molecules for such engine configurations [15].

Microalgae biodiesel is a very promising third-generation biofuel because of high lipid productivity and not competing with aliments resources. Microalgae biodiesel operates as a pilot fuel with high reactivity in PO-fuelled PCCI engines, offsetting the low cetane number of pine oil. Due to its high cetane number (generally >50) and long-chain saturated fatty acid profile, microalgae biodiesel can spontaneously ignite under the cool temperature and low-density conditions typical of PCCI operation [16]. The pilot injection strategy of microalgae biodiesel allows reactive stratification in the cylinder that facilitates controlled ignition of the premixed PO charge and hence lower normalized pressure rise rates and smooth combustion [17]. Beyond combustion benefits, fuels from microalgae results in a lower well-to-wheel carbon intensity well in line with wider sustainability aims [18].

A systematic approach has to be made in the optimisation of modern multi-fuel PCCI engines due to the large number of interacting control parameters (compression ratio, injection pressure, pilot–main fuel split, EGR rate, etc.) that are commonly used. Due to characteristics of nonlinear interactions between variables as well as the nature of local optimal solutions, the one-factor-at-a-time (OFAT) approach is not appropriate for such a complex system (Veza et al., 2023). (Lestari et al., 2024). Thus, recent research highlights a shift towards the design of experiments (DoE) and multi-objective optimisation frameworks to meet conflicting trade-offs

such as NO_x–soot emissions in modern engine calibration (Passerine & Breitkreitz, 2024; Ameen et al., 2023). A systematic approach has to be made in the optimisation of modern multi-fuel PCCI engines due to the large number of interacting control parameters (compression ratio, injection pressure, pilot–main fuel split, EGR rate, etc.) that are commonly used. Due to characteristics of nonlinear interactions between variables as well as the nature of local optimal solutions, the one-factor-at-a-time (OFAT) approach is not appropriate for such a complex system [19]. OFAT based experimentation is a discrete choice, it may be resource-heavy as well as it could skip the real optimum operation points. In addition, OFAT cannot be used in a predictive way because it does not include the noise of the experiments, e.g., unknown ambient conditions in the lab, which lead to unreliable results [20]. Thus, recent research highlights a shift towards the design of experiments (DoE) and multi-objective optimisation frameworks to meet conflicting trade-offs such as NO_x–soot emissions in modern engine calibration [21].

Response Surface Methodology (RSM) is often used in engine research for the modelling and optimisation of complex combustion processes. RSM uses structured experimental designs (e.g., Central Composite Design (CCD) and Box–Behnken Design (BBD)) to model system behaviour with quadratic polynomial models based on a small number of experiments. This allows one to quantify not only main effects but also interaction effects between input variables, the latter remaining relevant especially for spatio-temporally unstable combustion regimes like PCCI [22]. RSM has previously proven successful in optimising multi-fuel engines by finding optimal blend ratios and injection strategies that increase brake thermal efficiency along with a dramatic reduction in emissions [23]. The desirability function in RSM enables the multi-objective optimisation combining conflicting responses, e.g., NO_x and soot emission, into one objective function to perform optimisation [24]. But the very high R^2 (average $R^2 > 0.95$) reported in the literature suggests good prediction capability [25], RSM assumes a continuous response surface and is less suited for highly nonlinear or chaotic combustion phenomena [26] where complementary techniques, e.g., chaos control, need to be integrated.

Although statistical methods like RSM can be useful for local optimisation, the increasing complexity of modern combustion engines has led to increased interest in machine learning (ML) methods such as artificial neural networks, random forests and XGBoost. These data-driven models can express very nonlinear and high-dimensional relations between input and output; in addition, they do not need a predefined functional form, which makes these models extremely useful for complex prediction of emission behaviour in multi-fuel PCCI engines [27].

Nevertheless, the traditional black-box nature of ML models has hindered their application in practical engines; interpretability is crucial. Explainable artificial intelligence (XAI) particularly SHapley Additive exPlanations (SHAP) have been developed to address this limitation by quantifying the contribution of input features to model predictions in a game-theoretic context [28]. It has also been proven to be powerful to elucidate the major control parameters for a variety of operating conditions, such as the impact of injection pressure on soot formation or fuel blend ratio on NO_x emissions [29]. Due to the tightly coupled nature of PCCI engine systems, SHAP not only opens up compact, fine-grained insight into combustion behaviour compared to conventional global performance quantities but also provides further guidance on the development of robust and interpretable control strategies for advanced engines [30,31].

In recent years, machine learning (ML) and explainable AI (XAI) techniques have been explored with increasing interest for modelling and interpreting complex combustion and emission behaviour of internal combustion engine systems. For example, Chen et al. (2024) integrated ML–XAI into an existing fuel design framework for heavy-duty engines to improve interpretability and reliability of predictive models using a bridging approach between neural complexity and feature insight to construct new decision support tools through combustion optimisation [32]. Explainable machine learning to predict vehicle CO_2 emissions: The case of engine and operational inputs in automotive development used accumulated local effects, thereby showing how emissions change based on various engine and operational inputs, enabling development knowledge on the pathway to transport sustainability [33]. A hybrid ML model that combines ensemble and boosting algorithms with SHAP analysis was used to predict microalgae biodiesel engine performance and emissions [34]. Additionally, real-time fuel consumption and emission prediction using ML regression models with higher predictive accuracy was demonstrated. Such studies show the relevance of the interpretable ML frameworks that predict and explain rather than just predict complex engine phenomena and will provide the motivation for the integration of the two in the current work [35].

Concern about data integrity, repeatability and reliability in experimental research and development (R&D) increasingly led to the application of Total Quality Management (TQM) principles, which were pioneered in the manufacturing sector. With elaborate engine experiments, especially in unstable combustion modes such as PCCI and variable biofuel blends, systematic means to mitigate experimental variability and to control noise by the fuel preparation and ambient condition would help to facilitate the use of TQM frameworks, including

Quality by Design and Six Sigma [36]. Quality engineering tools (e.g., the Taguchi method) improve robustness by optimising signal-to-noise ratios and identifying operational conditions that are immune to uncontrolled factors [37,38]. More recent studies reconfirm this claim by introducing the term “Quality 4.0, which in turn combines the principles of TQM and the principles of Industry 4.0, such as machine learning, data analytics, etc., to allow early identification of experimental deviations and promote continuous improvement in a research context [39]. This kind of coupled framework guarantees that high-quality statistically robust experimental data will constitute the basis of advanced RSM and ML analyses [40].

The regular NO_x–soot trade-off in compression ignition engines has led to an increased interest in recent years in LTC, and PCCI in particular. Although PCCI emits less, it has a narrow stable operating range that requires fuels with specific volatility and reactivity. The use of volatile pine oil (PO) for it to act as a premixing enhancer and high-cetane microalgae biodiesel for its provision of consistent pilot ignition. On the other hand, the inherent interactions among key parameters (e.g., compression ratio, injection strategy, or blend composition) mean a one-factor-at-a-time approach does not help combust efficiently such fuel combinations.

Pairing these tools with proven experimental quality frameworks is less common, although RSM, machine learning (ML), and explainable AI tools, including SHAP, have each been used independently toward engine optimisation. Significantly, the PCCI engine has not yet systematically investigated the PO and microalgae biodiesel of the main–pilot fuel coupling ratio. In addition, the TQM-driven steps of combining RSM- and ML-based optimisation with robustness and interpretability validation of the experimental-optimal designs constitute an obvious research gap in the current PCCI combustion literature.

The primary focus of this study lies in the establishment of a unified effective framework for VCR PCCI engine performance and emissions enhancement. A few of the specific objectives are to:

- Experimental Characterisation: Investigation on Combustion, Performance and Emission Characteristics of VCR Engine Operating in PCCI Mode with PO as Main Fuel and Microalgae Biodiesel–Diesel Blends as Pilot Fuel
- Statistical Optimisation (RSM): To use Response Surface Methodology to obtain the optimal combination of engine parameters (compression ratio, injection pressure and EGR rate) and fuel blend ratios that achieve maximum thermal efficiency with minimum NOx and soot emissions.
- Predictive Modelling & Interpretability (ML & SHAP) → Build accurate machine learning models to predict engine outputs,

and use SHAP (Shapley Additive Explanations) to measure the contribution of each input parameter to emissions formation, revealing the physical mechanisms of emissions.

- The Robust Quality implementation (TQM): Implement TQM concepts, particularly Taguchi's Robust Design, so that engine performance is less influenced by experimental noise factors, thus ensuring reliable and reproducible engine operation.

2. Materials and Methodology

2.1 Materials and Test Fuels

All the fuels tested in the current study included one standard pilot fuel, two biodiesel–diesel pilot blends, and an oxygenated primary fuel for PCCI operation. Neat diesel (D100), microalgae biodiesel and PO were all purchased from a certified local commercial vendor. The microalgae biodiesel used in this study was derived from **Chlorella vulgaris**, a freshwater green microalga known for its high lipid content, favourable fatty acid profile, and widespread application in biodiesel research. All fuels in the range were subsequently pre-tested and characterised in the laboratory prior to the run-up to engine experimentation to ensure that they correspond to standard fuel quality requirements. Two pilot fuel blends were prepared by blending 10% relative volume of microalgae biodiesel (CV10, 10% biodiesel + 90% diesel) and 20% relative volume of microalgae biodiesel (CV20, 20% biodiesel + 80% diesel) with diesel. The premixed fuel used was PO with supply at 10%, 20% and 30% by volume, which are denoted as PO10%, PO20% and P30%, respectively. In the PCCI configuration, the chosen pilot fuel (D100, CV10 or CV20) was injected directly into the combustion chamber through the conventional fuel injection system, where it ignited, while PO was introduced by a port fuel injection arrangement at the intake manifold to create a homogeneous premixed mixture. This dual-fuel arrangement allowed low-temperature combustion to be controlled but had better premixing features.

Pine oil ranged from 10–30% v/v on the basis of preliminary trials and combustion stability under PCCI operation. Premixed combustion behavior was minimally affected by substitution levels less than 10%; However, due to the low cetane number and high volatility of pine oil, levels greater than 30% led to unsteady combustion with a high pressure rise rate and increased cycle-to-cycle variability. Thus, 10%, 20% and 30% substitution levels were selected as a rational and stable operating window for systematic investigation.

Table 1. Physicochemical Properties of Test Fuels

Property	Unit	Test Method	Diesel (D100)	Microalgae Biodiesel	CV10	CV20	PO
Density at 15 °C	kg/m³	ASTM D1298	**832**	**885**	**837**	**842**	**872**
Kinematic Viscosity at 40 °C	mm²/s	ASTM D445	**2.62**	**4.82**	**2.94**	**3.22**	**2.38**
Lower Heating Value (LHV)	MJ/kg	ASTM D240	**42.8**	**37.9**	**42.1**	**40.7**	**41.4**
Cetane Index	–	ASTM D976	**50.1**	**53.4**	**50.6**	**51.2**	**41.2**
Flash Point	°C	ASTM D93	**54**	**164**	**62**	**71**	**52**
Oxygen Content	wt.%	Elemental Analysis	**0**	**11.3**	**1.13**	**2.26**	**8.4**
Stoichiometric Air–Fuel Ratio	–	Calculated	**14.5**	**12.6**	**14.2**	**13.9**	**12.9**
Carbon Content	wt.%	Elemental Analysis	**85.1**	**73.4**	**83.9**	**82.7**	**76.5**

The density, kinematic viscosity, lower heating value (LHV), cetane index, flash point, and oxygen content of diesel, microalgae biodiesel, and CV10, CV20, and pine oil were measured by appropriate ASTM standards. Density, viscosity and LHV were measured by using a digital density meter, kinematic viscometer and bomb calorimeter, respectively. The cetane index was calculated via standard empirical correlations, while the flash point and the oxygen content were measured by a Pensky–Martens apparatus and an elemental analyser, respectively. Table 1 summarises the measured properties. Volume-weighted mixing rules were used to calculate effective properties of blended fuels, with these values applied to response surface modelling and machine learning analyses.

2.2 Experimental Setup

A detailed experiment has been performed on a single-cylinder, four-stroke, water-cooled, direct-injection variable compression ratio (VCR) diesel engine attached to a calibrated eddy current dynamometer for accurate load tracing. The engine was dual-fuel operation with pine oil as the premixed main fuel supplied via the intake manifold with port fuel

injection, while diesel, CV10, or CV20 was directly injected into the combustion chamber as pilot fuel. A tilting-head mechanism was used to vary the compression ratio. Cooling water temperature was kept within the manufacturer-recommended limits for all tests, and all tests with the engine were carried out at the rated engine speed. The brake power, fuel consumption, and brake thermal efficiency were calculated using the standard thermodynamic relations, and brake power was measured with a measuring dynamometer, while the fuel consumption was measured using a calibrated burette and electronic timer. The calibration of all the instrumentation was undertaken beforehand to guarantee accurate and repeatable measurements throughout the experimentation.

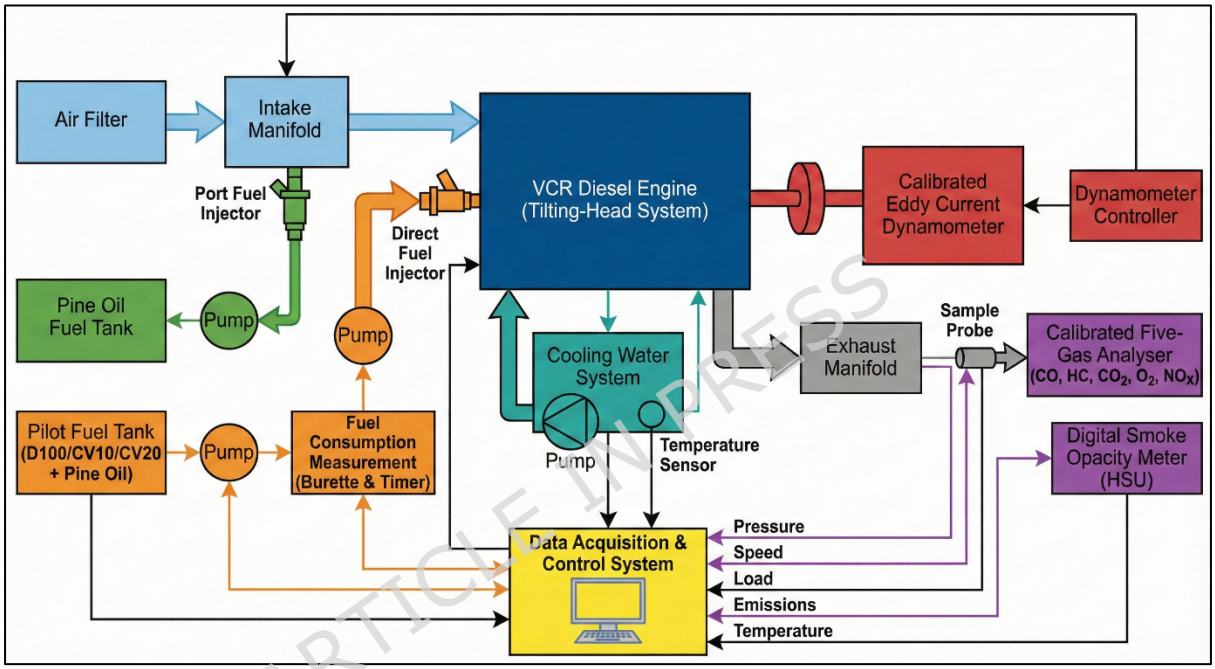


Figure 1. Experimental setup

Exhaust gas emissions were evaluated by a calibrated five-gas analyser. It was capable of measuring HC, total unburned hydrocarbons, carbon dioxide, oxygen (O_2), and NOx. Smoke opacity was monitored using a digital smoke opacity meter in Hartridge Smoke Units (HSU). To ensure proper mixing of the gases, the sample probe was placed in the exhaust manifold well upstream of the exhaust valve. Standard reference gases were employed to zero-calibrate and span-check both the gas analyser and the smoke meter prior to each experiment. The detailed experimental setup displayed in the table 2 and schematic diagram shown in figure 1.

Table.2 Experimental Setup

Engine Specification	
Parameter	**Specification**
Engine type	Single-cylinder, four-stroke, DI CI engine
Cooling system	Water-cooled

Aspiration	Naturally aspirated
Rated power	3.5 kW
Rated speed	1500 rpm
Bore × Stroke	87.5 mm × 110 mm
Compression ratio range	16:1 – 19:1 (variable)
Displacement volume	661 cm^3
Injection system	Mechanical direct injection
Injection timing	23° bTDC
Injection pressure	200 bar
Dynamometer	Eddy current type
Load measurement	Strain-gauge based load cell
Speed measurement	Magnetic pickup sensor
Starting	Electric start

Specifications of Emission Measurement Devices

Parameter	Measuring Range	Resolution	Accuracy
CO	0–10 % vol	0.01%	±0.02 %
HC	0–10,000 ppm	1 ppm	±10 ppm
CO_2	0–20 % vol	0.10%	±0.2 %
O_2	0–25 % vol	0.01%	±0.1 %
NOx	0–5,000 ppm	1 ppm	±10 ppm
Smoke opacity	0–100 HSU	0.1 HSU	±1 HSU
Exhaust gas temperature	0–900 °C	1 °C	±2 °C
Fuel flow rate	0–30 ml/min	0.1 ml	±0.5 %

2.3 Experimental Procedure

All the measuring equipment, including the fuel flow metering system, gas analyser, smoke meter, thermocouples and load cell, was calibrated against the manufacturers' instructions before commencement of the experiments. Initially the engine's cooling water circulation and lubricating systems were tested properly. Initially, the engine was operated at the baseline compression ratio on clean diesel (D100) to achieve quasi-steady-state temperature levels. Once warmed up, the fractional compression ratio (16, 17.5, or 19) was set using the VCR mechanism. The engine retained its stock mechanical direct-injection (DI) system for the pilot fuel (diesel, CV10, or CV20), which was injected directly into the combustion chamber at the specified injection timing and pressure. Pine oil (PO) was supplied separately as the premixed fuel through intake manifold injection using a dedicated port delivery line to establish PCCI operation. The pine oil substitution levels (PO10%, PO20%, and PO30%) were controlled by adjusting the volumetric flow rate of PO

relative to the total liquid fuel supplied (pilot fuel + PO), using calibrated fuel flow measurement devices. These substitution percentages were verified at each operating point under steady-state conditions by monitoring the individual flow rates of the pilot fuel and PO prior to data acquisition, ensuring that the target energy share and mass ratios were maintained within experimental tolerances. For every mixture of fuel, the engine was operated at a constant rated speed running under six separate load conditions (0%, 20%, 40%, 60%, 80%, and 100% of full load). These loads were applied using an eddy current dynamometer. Prior to each data collection, the engine was allowed 3–4 minutes to come to steady-state at various operating points. The steady state was confirmed when the engine speed, exhaust gas temperature, and fuel flow rate were constant over time.

Once steady state was achieved, the following behaviours were measured simultaneously:

- Engine speed and brake torque,
- Fuel consumption,
- Exhaust emissions (CO, HC, and NOx),
- Smoke opacity.

The measured torque and speed were used to calculate brake power, and brake thermal efficiency (BTE) and brake specific fuel consumption (BSFC) were estimated based on standard thermodynamic relations. Any drift of the instrument, as indicated by the calibrations, was corrected from the actual emission readings taken in this study. Repeat data in each test condition to allow experimental repeatability together with minimising random uncertainty was obtained by repeating each test condition four times, and arithmetic mean data of the measured values were obtained for further analysis. The overall experimental matrix included tests with three pilot fuels (D100, CV10 and CV20), three PO blends (10%, 20% and 30%), three compression ratios (16, 17.5 and 19) and six load conditions. The fuel supply lines were flushed with diesel to prevent cross-contamination between different fuel combinations after the completion of each test series. Further, all experimental runs were done keeping identical ambient conditions as far as possible. This full experimental data set acquired from these experiments was then used for response surface modelling, multi-objective optimisation, machine learning prediction, SHAP explainability analysis, and finally TQM uncertainty quantification.

2.4 Response Surface Methodology (RSM) Design and Mathematical Modelling

The combined effects of the compression ratio, engine load, and PO proportion on the performance and emission characteristics of a PCCI engine were then modelled and optimised using response surface methodology (RSM) by using **Minitab® statistical software**. RSM

combines the statistical design of experiments with the mathematical regression analysis that creates predictive mathematical models for the system at hand and evaluates the various optimum operating conditions with a minimum number of experimental runs. This study has chosen a Central Composite Design (CCD) to maximise its efficiency in estimating first-order, interaction, and quadratic effects of input parameters. The RSM study consisted of three factors, which were the compression ratio (CR), load of the engine (Load) and proportion of PO (Pine). The block factor for the pilot fuel type (D100, CV10, and CV20) mitigated systematic variation due to the differences in the pilot fuel and enhanced the confidence of the models developed.

The input parameters for the RSM investigation are described in terms of coded and uncoded amounts in Table 3. Ranges of the factors were set from the preliminary experiments, safety limits of the engine, and stable PCCI performance. An experiment was designed using a face-centered CCD. The design consisted of: 8 points of factorial, 6 axial points, and 6 center points, totaling 20 trial runs. The tests were divided into three blocks according to systematic differences between the pilot fuels (D100, CV10 and CV20).

Table 3. Coded and Uncoded Levels of RSM Input Factors and design matrix

Factor	Symbol	−1 (Low)	0 (Center)	+1 (High)
Compression Ratio (CR)	A	16	17.5	19
Engine Load (%)	B	0	50	100
PO Proportion (%)	C	10	20	30
Design Matrix		**CR**	**Engine Load (%)**	**PO Proportion (%)**
		19.0	50	20
		17.5	50	10
		17.5	50	20
		16.0	50	20
		17.5	0	20
		17.5	50	20
		17.5	100	20
		17.5	50	30
		16.0	0	30
		16.0	100	10
		17.5	50	20

	19.0	100	30
	19.0	0	10
	17.5	50	20
	19.0	0	30
	16.0	100	30
	19.0	100	10
	16.0	0	10
	17.5	50	20
	17.5	50	20

Table 4. Summary of CCD Used in the Present Study

Design Parameter	Value
Design type	Face-centered Central Composite Design
Number of factors	3
Number of blocks	3 (D100, CV10, CV20)
Number of factorial points	8
Number of axial points	6
Number of center points	6
Total number of experiments	20
α value	1 (face-centered)

The experiments were done systematically with the use of the CCD (Table 4) matrix that was generated through Minitab. BTE, BSFC, HC, unburned HC, nitrogen oxides (NOx) and smoke opacity were the outputs of the responses measured at each design point. For each response, we fitted a second-order polynomial regression based on the input variables and each of the output responses. Statistical validation of each developed RSM model (Analysis of Variance (ANOVA), Coefficient of determination (R^2), Adjusted R^2, Predicted R^2, Lack-of-fit test, and Residual diagnostic analysis (normal probability plot, residuals vs fitted, and residuals vs run order)) was done to establish the adequacy of each developed RSM model. The lack-of-fit p-value had to be higher than 0.05 for the developed model to be accepted as a good representation of the experimental data, and the model p-value was less than 0.05, indicating that the model was significant. Quadratic models showed good predictive ability with high R^2, adjusted R^2 and predicted R^2 values.

Once we had regression models that were statistically significant for all responses, the desirability function approach was implemented for multi-objective optimisation. The eight typical objective functions were taken into account to maximise BTE and minimise BSFC, CO, HC, NOx and smoke opacity. Additionally, the relative importance of the consensus

weights was then engineered into each response based on environmental relevance. The response ergonomiser of Minitab was used to identify the best level of compression ratio, load and PO fraction which maximises the overall desirability. This ideal regime, which was projected, then correlated well with machine-learning models and experimental trends.

2.5 Machine Learning Methodology, Model Validation, and GUI Development

Machine learning (ML) models to predict engine performance and emissions were developed using an experimental dataset and RSM design matrix (Python PyCharm), as the summarised flowchart of the process showed in Figure 2. Three input variables used were compression ratio, engine load, and pine oil (PO) proportion, and four input variables were pilot fuel type (D100, CV10, CV20), and outputs were BTE, BSFC, HC, CO, NO_x, and smoke opacity. Data used for all tasks are pre-processed using min–max normalisation, and categorical variables are numerically encoded, after which we use an 80:20 split for train test. As a part of this work, nine regression models (Gradient Boosting, XGBoost, Gaussian Process Regression, Extreme Learning Machine, Random Forest, Support Vector (SVR), LightGBM, Artificial Neural Network and Deep Neural Network) were implemented to capture the nonlinear behaviour of PCCI combustion. Hyperparameters were especially optimised via grid search using cross-validation. For comparing models, the R^2, MAE, MSE and RMSE metrics were used, and predicted-versus-actual plots were used to validate the models. SHAP analysis, which represented the effect of input parameters on predicted outputs, was used for model interpretability, which showed that BTE and NO_x were majorly influenced by engine load and compression ratio, while CO, HC and smoke were dominated by PO proportion. This paper presents a Python-based graphical user interface (Tkinter) to quickly predict responses to engine operating points that provide a robust digital decision-support tool to rapidly optimise the operation without repeated testing of the engine.

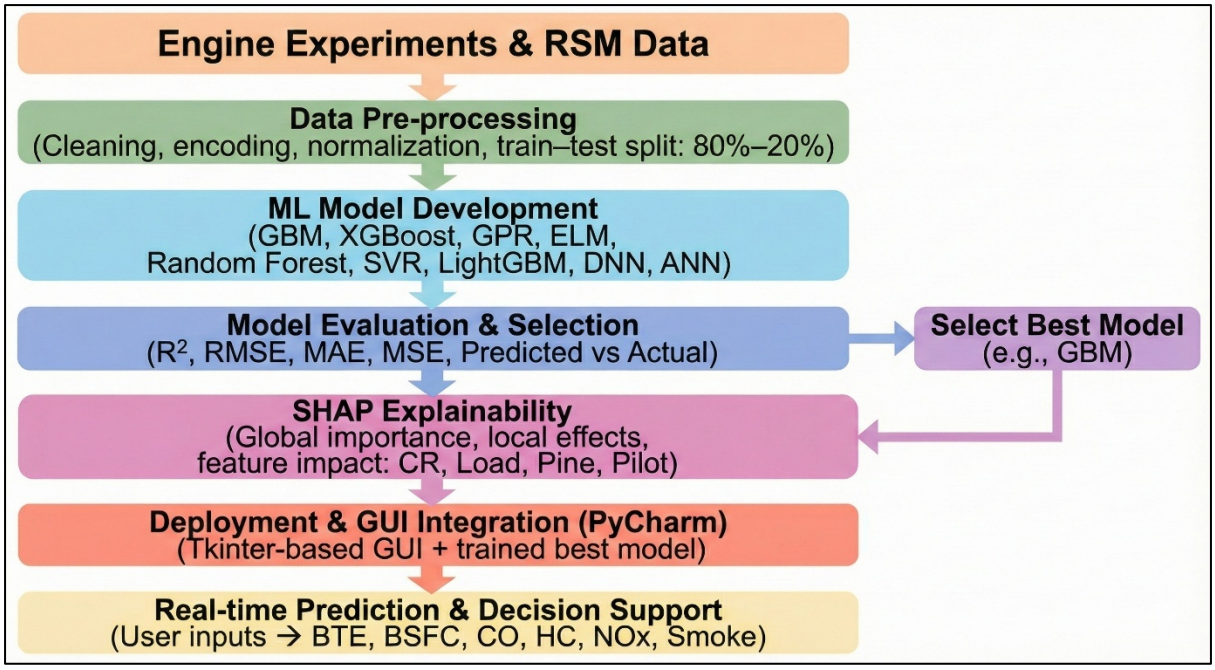


Figure 2. Machine Learning Methodology

2.6 Total Quality Management (TQM) Methodology

A Total Quality Management (TQM) framework based on the DMAIC (Define–Measure–Analyse–Improve–Control) methodology was integrated across the experimental, response surface methodology (RSM), and machine learning workflows to ensure data reliability, repeatability, statistical robustness, and continuous improvement of the PCCI engine investigation. Brake thermal efficiencies (BTEs), brake specific fuel consumptions (BSFC), HC, unburned HC, nitrogen oxides (NOx), and smoke opacity were defined as critical-to-quality (CTQ) parameters, while compression ratio, engine load, PO proportion, and pilot fuel type (D100, CV10, and CV20) were using quantitative statistical metrics rather than qualitative judgment alone as key process input variables. Set A: Experimental data were obtained with triple repetitions being averaged at each operating point to ensure repeatability, and the reliability of all measures was quantified by uncertainty propagation according to the root-sum-square method; the uncertainty of the resultant was evaluated as

$$U_R = \sqrt{\sum_{i=1}^{n} U_i^2} \quad (1) \quad [41]$$

where U_R is the resultant uncertainty and U_i are individual instrument uncertainties (fuel flow rate, load, emissions, etc.). The propagated uncertainty of derived parameters such as BTE and BSFC was calculated using

$$\frac{U_y}{y} = \sqrt{\sum \left(\frac{\partial y}{\partial x_i} \cdot \frac{U_{x_i}}{x_i} \right)^2} \quad (2) \quad [41]$$

where σis the standard deviation of the quality response. Based on this analysis, the overall experimental uncertainty was maintained within ±2.1%, which is well within acceptable limits for IC engine experimentation. Statistical quality analysis was carried out using ANOVA from the developed RSM models, residual diagnostics, Pareto analysis, and dispersion analysis using the standard deviation expression (σ)

$$\sigma = \sqrt{\frac{1}{n-1}\sum(x_i - \dot{x})^2}, \quad (3)\ [42]$$

which showed that BTE and NOx were most significantly affected by engine load and compression ratio while CO, HC, and smoke emissions were mainly affected by PO proportion. Desirability-based multi-objective optimization with RSM, aided by machine learning–based refinement and SHAP interpretability, identified the optimized combustion regime (CR = 18.94, load = 43.43%, PO = 30%), one that exhibited concurrent improvements in thermal efficiency and reductions in regulated emissions. The percentage improvement index was used to quantify the degree of improvement

$$PI = \frac{Y_{opt} - Y_{base}}{Y_{base}} \times 100 \quad (4)\ [42]$$

where $Y_{base}$and Y_{opt}represent baseline diesel and optimized values, respectively. Long-term quality control was ensured using statistical process capability analysis through the indices

$$C_p = \frac{USL-LSL}{6\sigma} \quad (5)\ [42]$$

$$C_{pk} = \min\left(\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\right) \quad (6)\ [42]$$

where USL and LSL are the upper and lower specification limits, and μis the process mean and all the optimized responses displayed Cpk > 1.33 indicates that the combustion process is statistically stable and capable. Finally, a set of key performance indicators was used monitor the overall quality performance representative of the closed-loop TQM in sustaining high-quality PCCI engine operation: (1) improvement of BTE reduction, (2) reduction of BSFC, (3) reduction indices in emission, (4) machine learning prediction accuracy, (5) repeatability as well as reproducibility of the experimental procedure, and (6) process capability.

2.7 Sustainability Assessment Methodology Using the Pugh Matrix

To evaluate the sustainability performance of the PCCI engine utilising diesel, CV10, CV20, and a pine-oil assisted combination, a systematic multi-criteria decision-making method, specifically the Pugh matrix, has been developed as illustrated in Figure 3. Key techno-environmental indicators were consolidated into a single comparative decision-making framework. A comprehensive comparative framework integrating various techno-environmental indicators The baseline established in this investigation was conventional diesel operation, serving

as the reference point for the evaluation of all alternative fuel combinations. Evaluation criteria were established that delineated energy efficiency, environmental impact, fuel characteristics, and operational attributes. This encompasses, specifically, brake thermal efficiency, brake specific fuel consumption, carbon monoxide, unburned hydrocarbons, nitrogen oxides, smoke opacity, fuel oxygen content, renewable fraction, and combustion stability. In the Pugh matrix analysis, uniform weights were allocated to each criterion to ensure objectivity and transparency throughout the decision-making process. The sustainability criteria were subsequently organised into one of four distinct categories: performance (BTE, BSFC), environmental (CO, HC, NOx, smoke), energy and climate (renewability, carbon footprint), and practical deployment (fuel availability, engine compatibility, safety and storage).

The Pugh matrix was evaluated qualitatively (better than, equal to, or worse than the baseline condition) utilising symbols (+, 0, −) that were subsequently converted to an aggregate character through numeric scoring (+1, 0, −1). To further assess the robustness of the sustainability rankings, a sensitivity analysis was conducted by varying the weights of the two primary criteria (BTE and NOx) by ±20%. Notably, when a moderate adjustment was made to the weight assignment, the overall ranking did not alter, with the optimised PCCI operating condition consistently surpassing the baseline and other fuel blends. This observation reinforces the notion that the decision framework exhibits resilience to slight variations in rank ordering. Throughout the implementation phase, the sustainability ranking derived from the Pugh method was further validated at an experimental level by correlating the sustainability of each design with the results from RSM optimisation and machine-learning predictions. This approach ensured that the sustainability decisions were supported by both experimental and statistical evidence.

The combined Pugh–RSM–ML framework enabled an optimal identification of the sustainable PCCI operation condition that maximised the performance, fuel economy, emissions and renewable fuel consumption objectives whilst ensuring combustion stability and system reliability. Performance tests on a VCR diesel engine run in premixed charge compression ignition (PCCI) mode using a single-cylinder, four-stroke, water-cooled engine. The blended fuels were primarily PO and diesel-based as pilot fuels. We purchased neat diesel (D100), CV10 (10% microalgae biodiesel + 90% diesel), CV20 (20% microalgae biodiesel + 80% diesel), and PO from a local retail store and conducted laboratory analyses in accordance with ASTM standards. The PO was injected via the port at the proportions of 10%, 20% and 30%, while the selected pilot fuel was injected into the combustion chamber directly. The engine was tested

from 0 per cent load to full load at six different points coupled with 16, 17.5, and 19 compression ratios. Performance and emissions were measured using a calibrated eddy current dynamometer, five-gas analyser, and smoke meter. Central composite design from Minitab was utilised to characterise and optimise the total effects of load, compression ratio, and the wt% of PO. In PyCharm, the models were created using nine regression algorithms to predict the performance and emissions of the machine learning models. Model interpretability was enabled through SHAP analysis, and a real-time GUI was developed for virtual prediction. For this purpose, a new DMAIC-based TQM framework adapted with the Pugh matrix method is proposed to control uncertainty and assess sustainability.

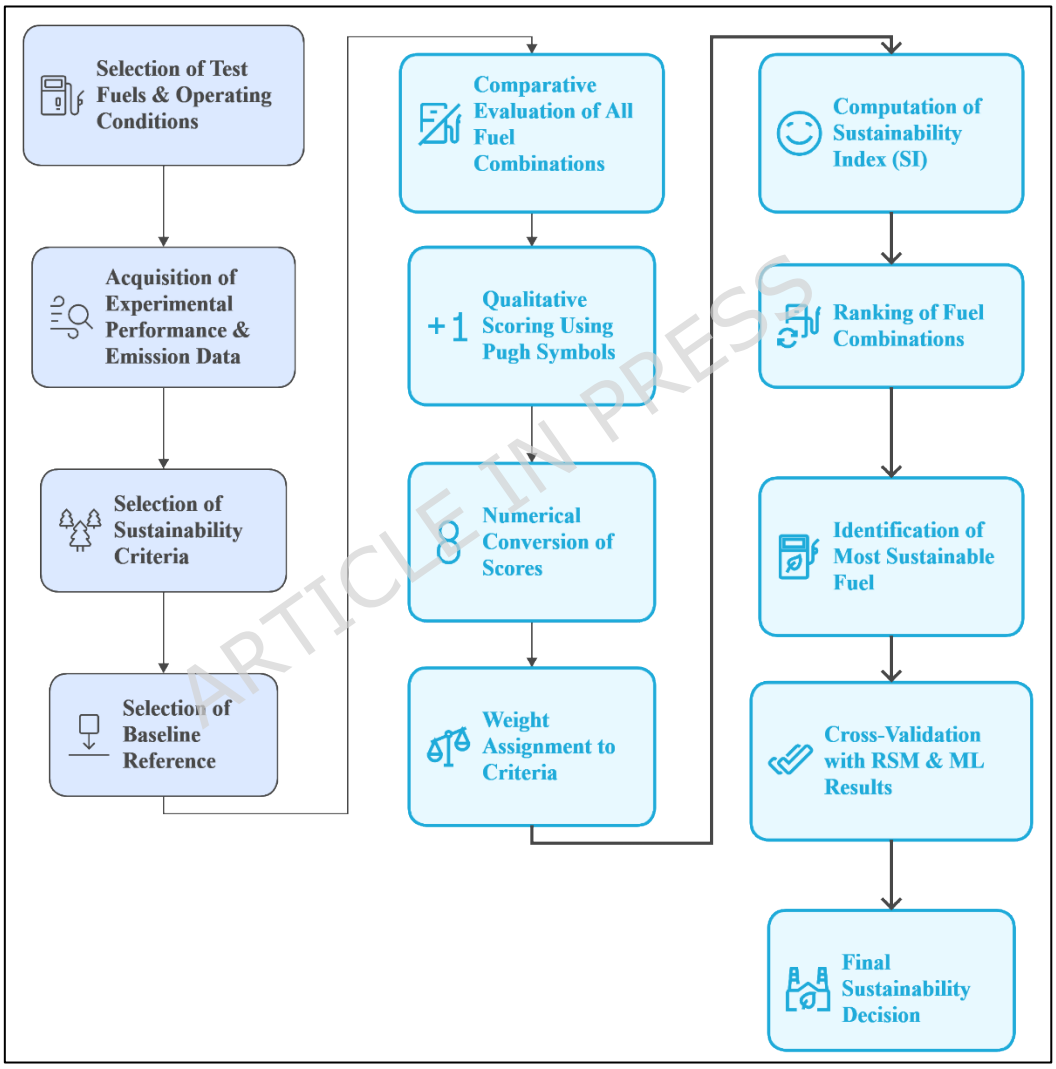


Figure 3. Pugh Matrix Methodology

2.8 Code Availability

The custom Python code used for data preprocessing, machine learning model development, hyperparameter optimisation, validation, explainability analysis (SHAP), and graphical user interface generation is provided as Supplementary (S3) Information. The code enables reproduction of all machine learning results reported in this study. No

proprietary software was used, and the code can be executed using standard open-source Python libraries.

3. Result and discussion

3.1 Combined Experimental and Response Surface Methodology (RSM) Analysis of Engine Performance and Emission Characteristics

3.1.1 Influence of Compression Ratio, Load, and PO Proportion on Brake Thermal Efficiency under PCCI Operation

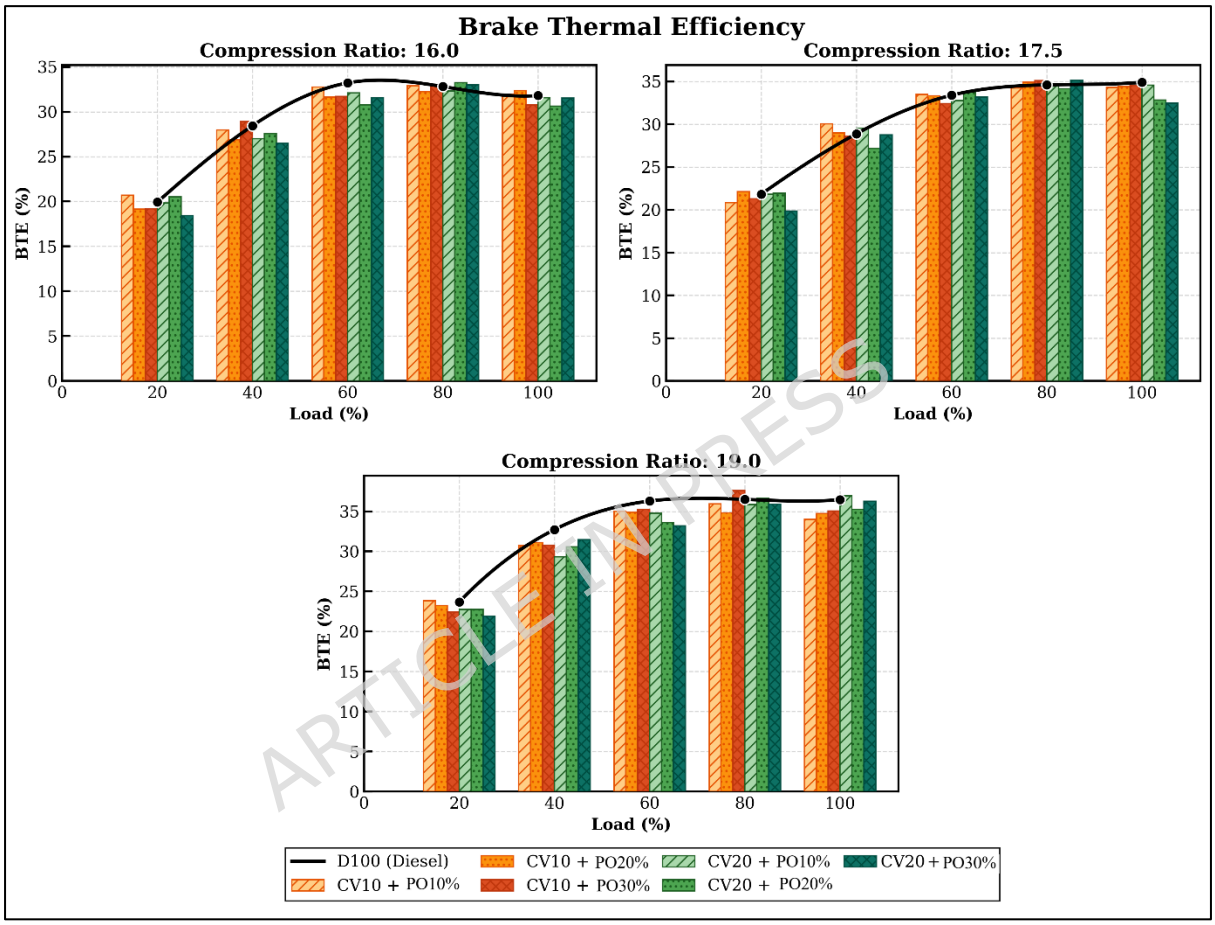


Figure 4. Effect of engine load on brake thermal efficiency under PCCI mode

When the pilot fuel was matched with different proportions of PCCI with D100, CV10, and CV20 in use, the variation of BTE with engine load under the influence of different compression ratios and quantities of PO is shown in Figure 4. BTE initially increases with the engine load from no load, reaching the medium-high load range (60–80%) for all fuel combinations. This is due to increased combustion efficiency, better fuel energy utilisation and decreased relative heat losses. For full load, not causing BTE to drop at saturation or a little, since the mixture is richer, so the exhaust losses are larger, and oxidation is not completely completed with rich fuel in the zone. The maximum measured value at 80% load has

been around 32.8–33.2% for the neat diesel (D100) at a compression ratio (CR) of 16. The maximum BTE increases to 36.0–36.5% at CR = 19, indicating that the compression ratio has been increased by almost 9–10%. The addition of PO affects BTE. The BTE increase changed from 0.82% for D100 to 0.97% for CV10 +P10%, 1.50% for CV10 +P20%, and close to 1.80% for CV10 +P30% from 33.39% to 35.10% at CR = 17.5, respectively, under 60% load. This represents an improvement of about 5% over using pure diesel. This is primarily because PO is extremely volatile and has an oxygenated property that helps in better air fuel mixing and faster premixed combustion phase. However, at low loads (0 to 20%), a higher amount of PO (P30) BTE decreases slightly due to the longer time required for the engine to start, and the possibility of partial misfire occurs due to a low cetane index of PO in cold weather. As it can be seen from Figure, CV10 has an essentially better BTE than D100 and CV20 in most of the operating region. The maximum experimental BTE results for CV10 +P30% are in the ranges of 36.5%–37.6% and 36.51% for D100, compared to 36.6%–36.9% for CV20 +P20% at CR = 19 and 80% load. It demonstrates that CV10, along with discrete and moderate-to-high PO substitution, provides optimal thermal efficiency performance. CV20 has a slightly worse BTE at high loads due to its required higher mass fuel flow rate and lower lower-heating value.

$$\begin{aligned} \text{BTE (\%)} = {} & -33.4 + 3.19\,CR + 0.7766\,Load - 0.035\,Pine \\ & - 0.0554\,CR\times CR - 0.004912\,Load \\ & \times Load + 0.00310\,Pine\times Pine + 0.00078\,CR\times Load \\ & - 0.00458\,CR\times Pine - 0.000463\,Load\times Pine \end{aligned} \quad (8)$$

The obtained quadratic regression model for BTE built using RSM is given in Equation (8). BTE was a function of compression ratio, engine load, and percentage of PO through linear, interaction, and quadratic terms. This model fits exceptionally well, boasting a coefficient of determination equal to or exceeding $R^2 = 0.9998$. It denotes the closeness of the predicted values to the actual values. Figure 5(a): Three-dimensional response surface of BTE as a function of (compression ratio) and (load). The corresponding contour plot is shown in figure 5(b). These graphs illustrate that the most influential factor that impacts BTE is engine load, followed closely by compression ratio. This is the area where the BTE reaches the highest value and is located at high compression ratios and moderate-to-high engine loads. The 3D response surface of BTE in terms of the amount of PO and the amount of load is shown in Figure 5(c). The results indicate that BTE increases with the addition of PO until the peak value within the studied range is reached, and this is more pronounced at the medium load.

The load versus compression ratio response surfaces presented in Figures 5 indicate that the interaction between both factors is statistically

significant. Which means that the positive effect of a higher compression ratio is at its great. This implies that a higher compression ratio exerts the greatest positive effect at moderate to elevated loads. As can be seen from Figures, the experimental data points closely fit the surface predicted by RSM, suggesting that the quadratic model is robust and reliable. The results of experimental and RSM analysis indicate that the optimal thermal efficiency of PO-assisted PCCI operation occurs at high compression ratios, moderate engine loads and moderately high proportions of PO. This is due to the combined relative benefits of improved thermodynamic efficiency, improved combustion phasing and improved pre-mixing characteristics.

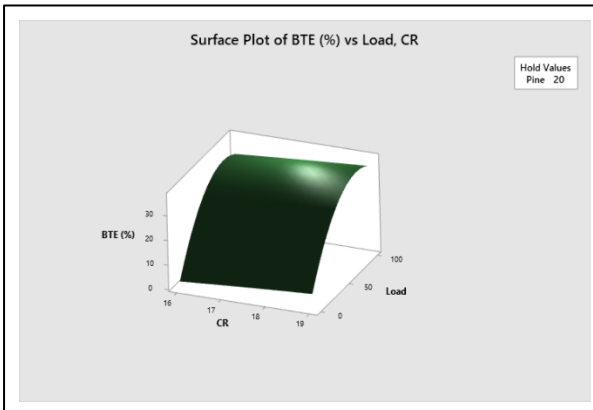


Figure 5 (a)

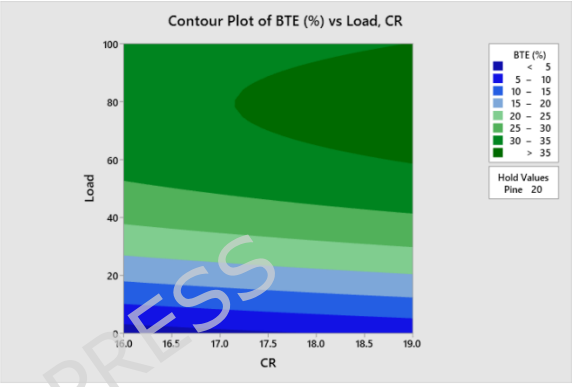


Figure 5 (b)

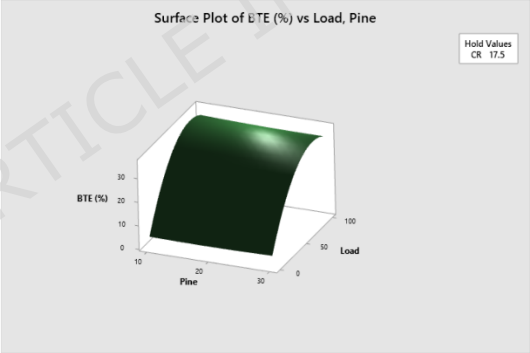


Figure 5 (c)

Figure 5. Response surface and contour plots of brake thermal efficiency: (a) effect of compression ratio and engine load, (b) contour map of compression ratio and engine load, and (c) effect of PO proportion and engine load.

3.1.2 Influence of Compression Ratio, Load, and PO Proportion on Brake Specific Fuel Consumption (BSFC)

The change in BSFC with engine load under operating conditions of PCCI for different compression ratios and proportions of PO are represented in Figure 6, where D100, CV10 and CV20 act as the pilot fuels. For all the fuel combinations, the BSFC shows typical behaviour of

decreasing to the minimum value and then increasing with load. BSFC at light load (0–20%) is rather high due to a large number of mechanical frictional and pumping losses and not using up all of the energy in the fuel supplied. When the load rises to medium-load levels (40–80%), the BSFC drops steeply as the combustion process is optimised and more chemical energy transforms into useful brake power. If the load is full, the BSFC is slightly higher because of rich mixture formation, higher heat transfer losses and not enough O_2 for combustion.

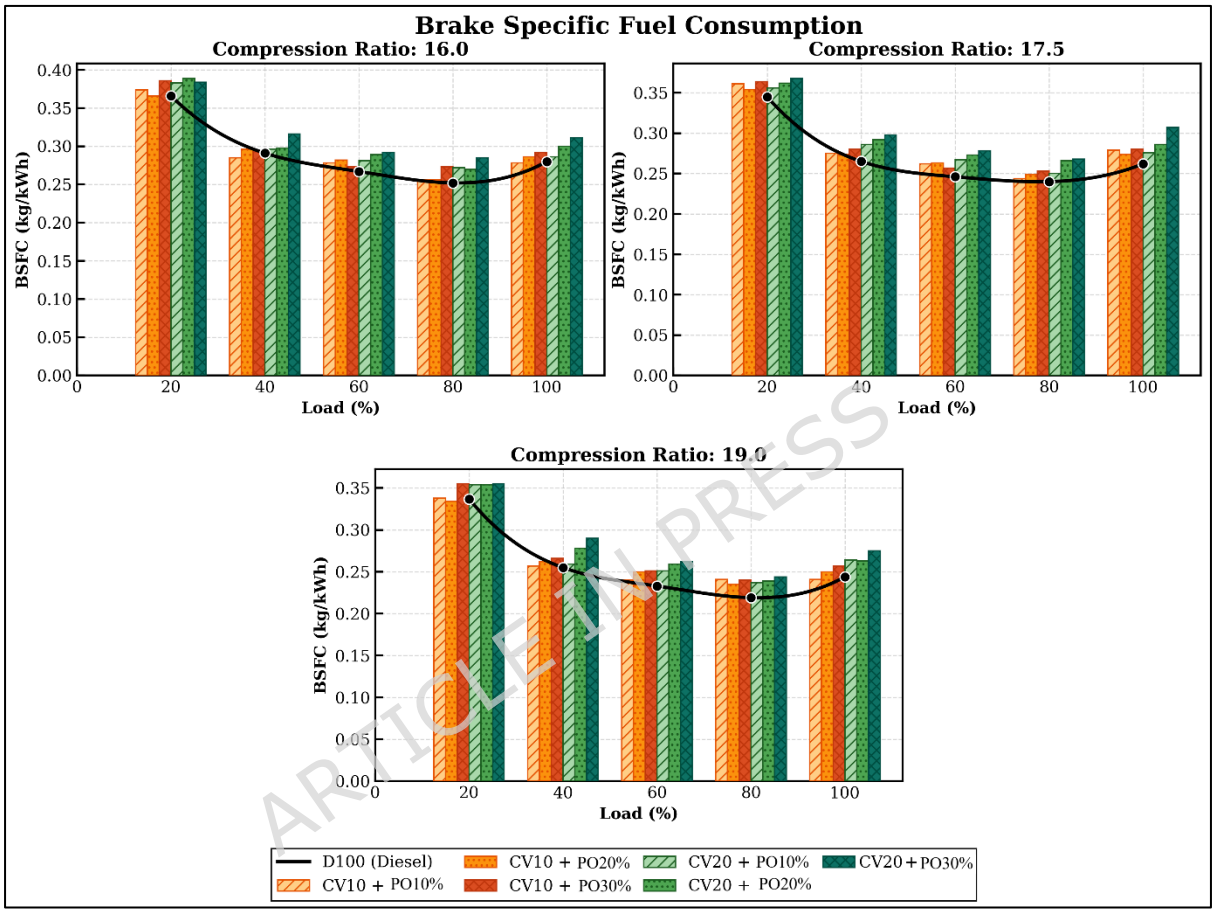


Figure 6. Effect of engine load on brake specific fuel consumption under PCCI mode

BSFC has fallen to about 0.366 kg/kWh at the 20% load and to about 0.252 kg/kWh at the 80% load with clean diesel (D100) and CR=16. Then it increases to approximately 0.280 kg/kWh at full load. At CR 19, the minimum BSFC decreases to 0.219–0.233 kg/kWh at 80% load. This means that change in compression ratio alone accounts for this 8-12% drop. The reason for this reduction is that at higher compression ratios, the thermal efficiency of the combustion is improved, and faster combustion over a shorter duration is observed. The effect of the addition of PO on BSFC. At CR = 17.5 and 60% load, BSFC decreases from 0.246 kg/kWh for D100 to 0.262 kg/kWh for CV10 +P10%, 0.263 kg/kWh for CV10 +P20% and 0.257 kg/kWh for CV10 +P30%. This indicates that a moderate quantity of PO has no detrimental effects on fuel economy for PCCI-mode operation of

the engine. At high loads the BSFC of all the PO blends is somewhat higher than D100 (80–100%) because the calorific values of the blends of PO and biodiesel are lower than that of petroleum diesel. However, this slight rise is negated by the rise in brake thermal efficiency in PCCI mode. CV10 generally has a lower BSFC than CV20 across the entire operating range of the pilot fuels, especially at medium-to-high load. Specifically, for a CR of 19 and a load of 80%, the minimum BSFC values are within the ranges of 0.240 kg/kWh for CV10+P30% and 0.239 kg/kWh to 0.244 kg/kWh for CV20+P20% and P30, respectively. In the same conditions, however, D100 produces approximately 0.219 kg/kWh. The BSFC of CV20 is slightly higher than that of CV10 due to the requirement of higher fuel mass flow and the lower lower-heating value to produce the same brake power levels.

$$\begin{aligned} \text{BSFC (kg/kWh)} = {} & 1.392 + 0.0288\,CR - 0.039989\,Load + 0.00330\,Pine \\ & - 0.001085\,CR\times CR + 0.000268\,Load\times Load \\ & - 0.000012\,Pine\times Pine + 0.000010\,CR\times Load \\ & - 0.000050\,CR\times Pine - 0.000017\,Load\times Pine \end{aligned} \quad (9)$$

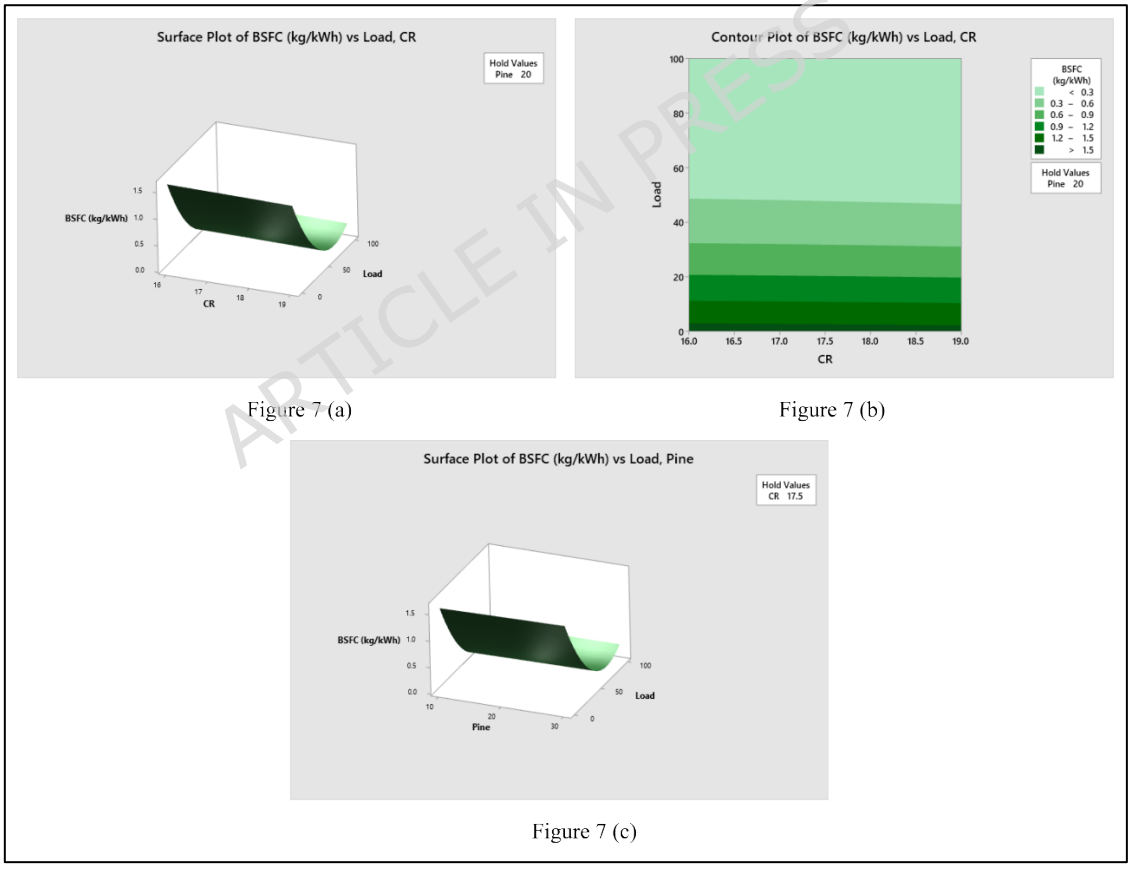


Figure 7 (a)

Figure 7 (b)

Figure 7 (c)

Figure 7. Response surface and contour plots of brake specific fuel consumption: (a) effect of compression ratio and engine load, (b) contour map of compression ratio and engine load, and (c) effect of PO proportion and engine load.

Using Eq. (13), the results of the quadratic regression model for BSFC are as follows: The predictive power displayed by the model is very

good; that is, The coefficient of determination is greater than $R^2 > 0.99$. That is, all the experimental BSFC lies very close to the value predicted from the equations. The 3D response surface of BSFC as a function of compression ratio and load is shown in Figure 7(a). The associated contour plot is shown in Figure 7(b). It is evident from these plots that BSFC decreases steeply with each increase in load and compression ratio. The minimum region is broad at high compression ratios and at medium-to-high engine loads. The response surface of BSFC with respect to the amount of PO and load is shown in Figure 7(c). It reveals that at high loads, BSFC increases slightly with a rise in the amount of PO due to the low calorific value of PO. At moderate work the level of pine oil expands; be that as it may, this does not influence an excessive amount of.

Figures 7 shows that the interaction between compression ratio and load is statistically significant in these two blocks, and the distinct curvature suggests that the beneficial effect of increasing compression ratio to reduce BSFC is most pronounced at moderate to high load [43]. The near overlap of the experimental data points with the RSM-predicted surfaces in Figures 7(a–c) confirms the robustness and reliability of the developed quadratic BSFC model. The results of experiments and RSM analysis indicate that the minimum fuel consumption is obtained at a high compression ratio, medium to high engine load and a moderate amount of PO. The reason being because it will have a higher combustion efficiency and better thermodynamic performance.

3.1.3 Influence of Compression Ratio, Load, and PO Proportion on Hydrocarbon (HC) Emissions

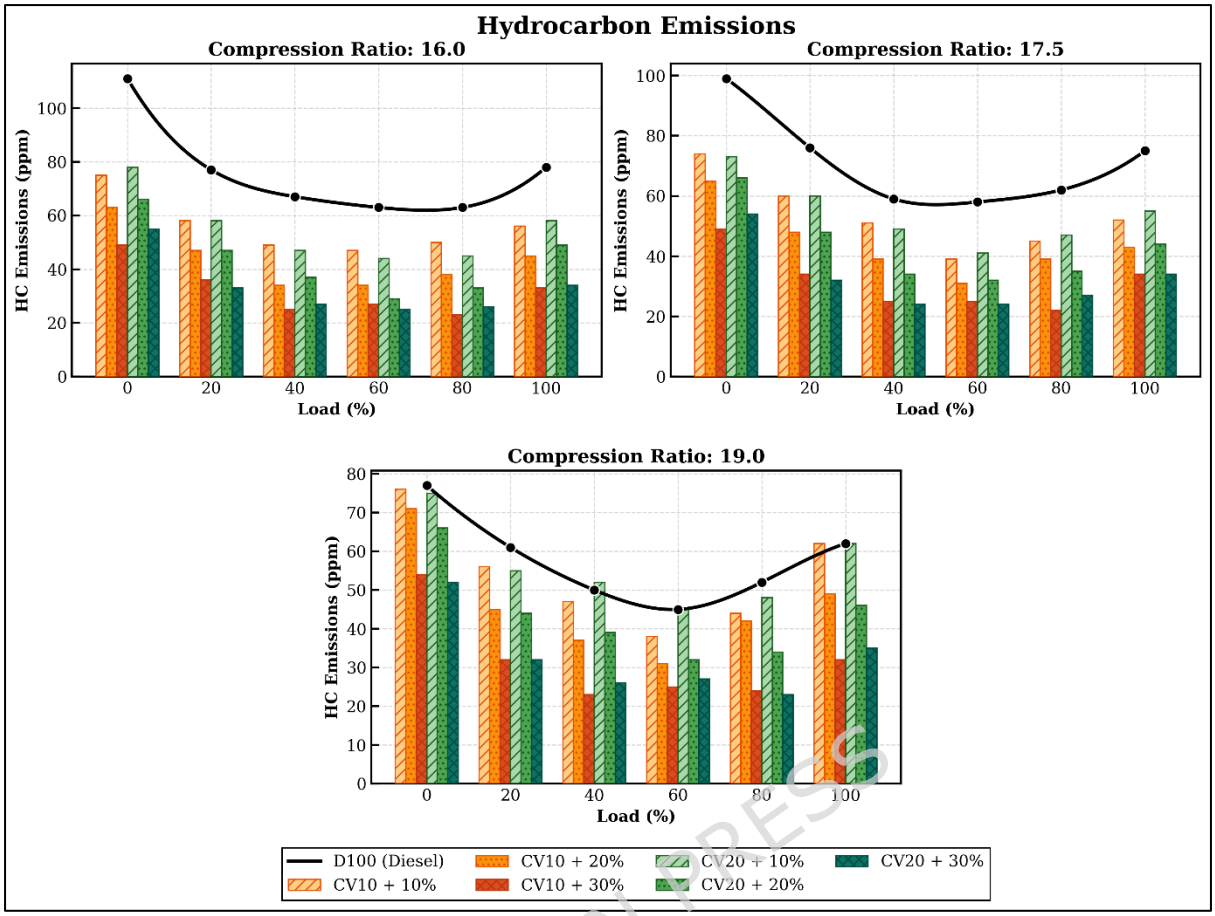


Figure 8. Effect of engine load on hydrocarbon emissions under PCCI mode

Unburned HC emissions versus engine load for various compression ratios and PO quantities when the engine is operated in PCCI mode with pilot fuels: (a) D100, (b) CV10, and (c) CV20 as pilot fuels shown in figure 8. VCG emissions are maximised at no load or low load for all fuel combinations. They get disabled one by one until the medium-load region (40–60%) is reached, as the load increases. Low temperature in the cylinder, slow oxidation kinetics and partial misfire at lower load also result in incomplete combustion and hence more HC formation. A higher HC emission at full load is observed due to localised fuel-rich combustion and lower availability of oxygen. Neat diesel (D100) leads to a decrease in HC emissions from 111 ppm at 0% load to 63 ppm at 60% load. And at 100% load, they back up again at almost 78 ppm. The HC values are greatly reduced when the compression ratio increases to CR = 19, decreasing to a maximum of approximately 77 ppm at no load and a minimum of 45 ppm at 60% load. It indicates that the levels are decreased by 25–30% due to the fact that the temperature in the cylinder increases and oxidation is better at high compression ratios.

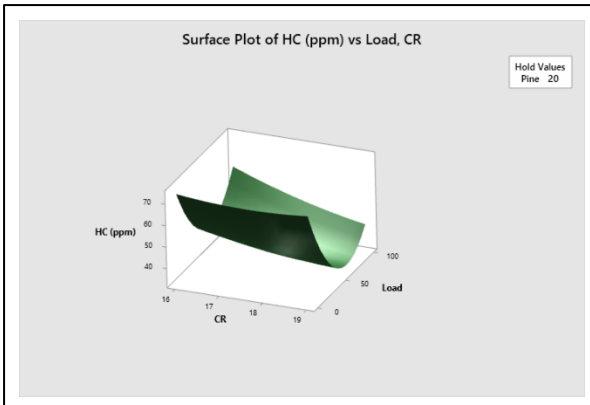


Figure 9 (a)

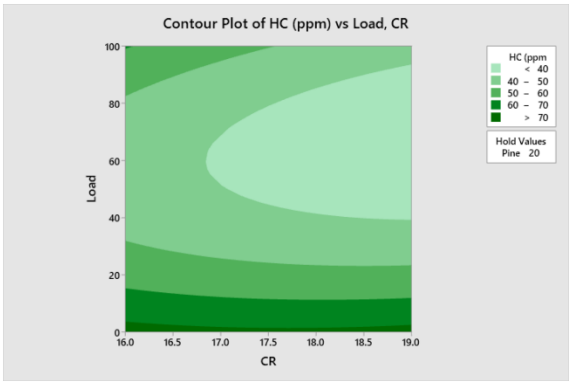


Figure 9 (b)

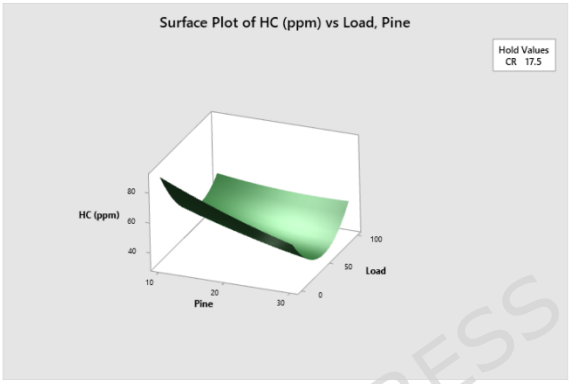


Figure 9 (c)

Figure 9. Response surface and contour plots of hydrocarbon emissions: (a) effect of compression ratio and engine load, (b) contour map of compression ratio and engine load, and (c) effect of PO proportion and engine load.

Aggressive cover with detection clear effect on HC emission caused by natural PO addition in figure 9. For example, at CR = 17.5 and 60% load, HC emissions decrease from 58 ppm for D100 to 49 ppm for CV10 +P10%, 39 ppm for CV10 +P20%, and then on the order of 25 ppm for CV10 +P30%. This is a nearly 57% maximum reduction compared with operation on pure diesel. The dramatic reduction in HC emissions comes from the fact that PO is highly volatile and has a high oxygen content. This facilitates premixed charge oxidation and prevents incomplete combustion from occurring. However, under very low load (0–20%), excessive addition of PO (PO30%) leads to slight enhancement of HC emissions due to the delayed ignition timing and poor flame propagation at low temperature PCCI condition. Under most operating conditions, CV10 always emits lower HC emissions compared with D100, and amongst the pilot fuels, CV20 has the highest HC emissions (except D100 at lower load conditions without the oxygenated pilot). The corresponding HC emissions are about 45 ppm for D100, 31-32 ppm for CV10 +P20%, and 32-34 ppm for CV20+P20% at a CR of 19 and a load of 60%. This indicates that CV10 is more suitable against PCCI operation. HC emissions are slightly elevated

with CV20 because higher biodiesel blends have greater viscosity and don't atomize quite as well, yielding a slightly worse local combustion quality.

Regression Equation in Uncoded Units

$$\begin{aligned} \text{HC (ppm)} = {} & 375 - 28.2\,CR - 0.414\,Load - 3.94\,Pine + 0.75\,CR\times CR \\ & + 0.00897\,Load\times Load + 0.0193\,Pine\times Pine \\ & - 0.0550\,CR\times Load + 0.092\,CR\times Pine \\ & + 0.01325\,Load\times Pine \end{aligned} \quad (10)$$

Eq. (10). It has a determination coefficient, i.e., $R^2 = 0.9717$, thus giving us accurate predictions. As illustrated in Figure 9(a), the three-dimensional response surface of HC emissions with respect to compression ratio and load. And its contour plot is shown in Figure 9(b). These graphs show more clearly that towards medium load, HC emissions decrease sharply with increasing compression ratio and load, respectively. The response surface of HC emissions as a function of the amount of pine oil and the load can be seen in Figure 9(c). The experimental results also revealed that the HC emissions are significantly reduced with the increase of PO, particularly under low-moderate engine loading.

3.1.4 Combined Experimental and RSM Analysis of CO Emissions

PCCI operation with different pilot fuel ratios using D100, CV10, and CV20 has also been examined in terms of CO emissions with respect to engine load as depicted in figure 10. At no-load and low-load, the CO emissions are relatively high in comparison amongst the tested fuel combinations. They drop sharply as you go into the medium-load range (40–60%) and then rise again slightly at full load. Higher CO emissions at low loads can be explained using the low in-cylinder temperature, slow oxidation kinetics and unburnt fuel caused at lean and instable PCCI conditions. The fall at mid-high loads is due to effective air/fuel mixing, higher combustion temperature and enhanced oxidation of CO to CO_2. The increase takes place at high load because of localised fuel-rich areas and decreases in oxygen concentrations, and because of quenching results next to combustion chamber surfaces. For neat diesel (D100) and a CR of 16, CO emissions decrease from approximately 0.077% at zero load to 0.037% at 40% load, then reach a minimum of approximately 0.035–0.037% and then increase to nearly 0.068% at 100% load. With a compression ratio of CR = 19, CO goes down further to about 0.048% at no load and to a minimum of about 0.020-0.025% around 40-60% load before going back up again to about 0.051% at full load. This results in a significant reduction in total CO in the order of 30–40% at higher compression ratios. The main reason is that at higher CR, the temperature in the cylinder is high, the oxidation rate of CO is fast, and the combustion time is short.

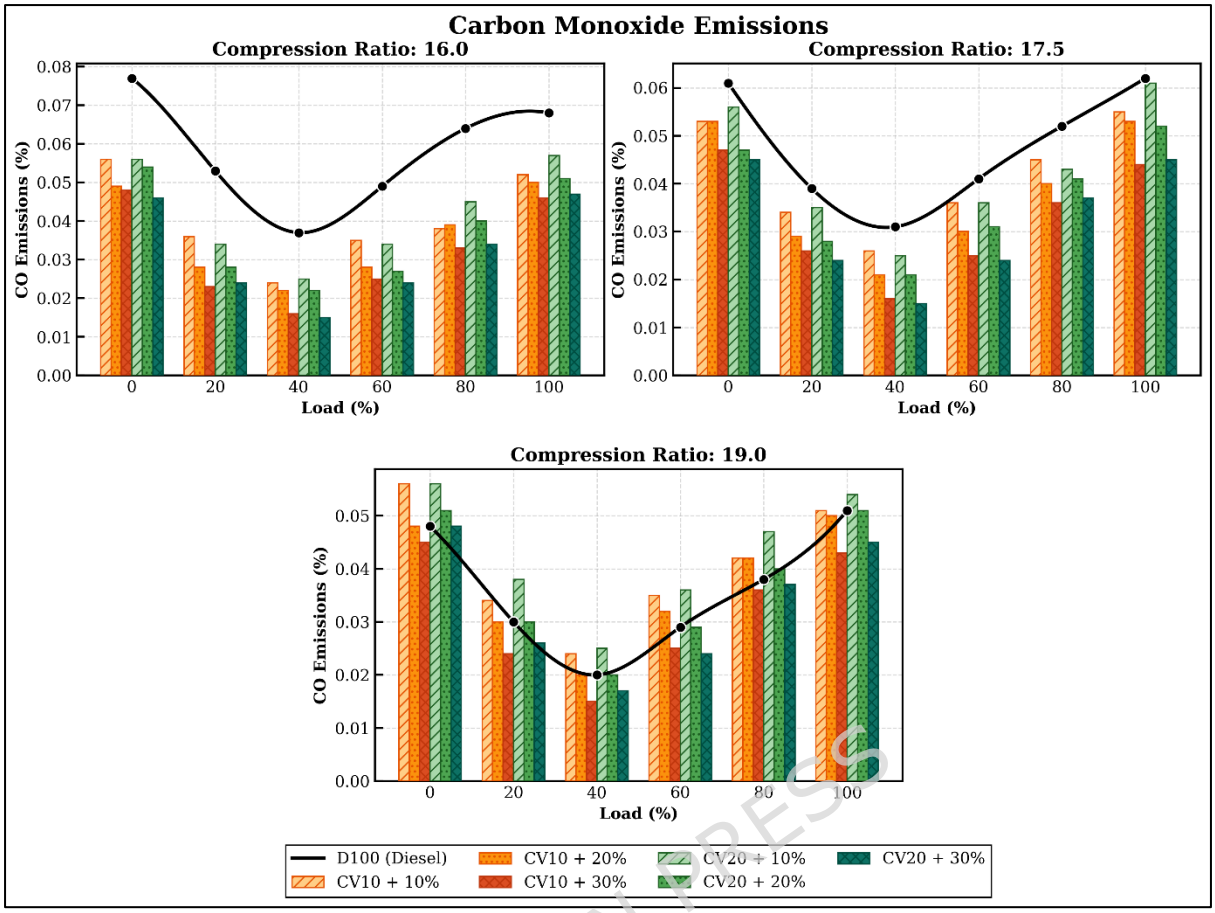


Figure 10. Effect of engine load on carbon monoxide emissions under PCCI mode

At 60% load and a CR of 17.5, CO emissions are reduced from approximately 0.041% for D100 to 0.030% for CV10 +P10%, 0.026% for CV10 +P20%, and to a minimum of approximately 0.025% for CV10 +P30%. This is a 39–42% maximum reduction from running neat diesel. This is mostly because PO has lots of oxygen, it's not particularly thick, it's highly volatile, and it combusts better with other fuels. To do so, this makes CO be oxidised quicker in the expansion stroke. However, when the load is very low (0–20%), the more pine oil imparted (PO30%) results in a slightly higher CO level than that of PO10% and PO20%. In low-temperature PCCI conditions, the ignition delay is long and the flame partially goes out; hence, NOx emission is less in those conditions. Under all operating conditions, CV10 always gives the lowest CO emissions of all the pilot fuels. At CR=19 and 60% load, CO emissions result in about 0.029% for D100, 0.025–0.026% for CV10 +P20%, and 0.029–0.031% for CV20 +P20%. This indicates that CV10 is more effective at PCCI operation due to its better oxidation and better spray interaction. This higher viscosity and density produce a slight increase in CO emissions with CV20 as it becomes more difficult to atomise and create locally rich mixtures. This, in turn, increases the amount of unburnt combustion products.

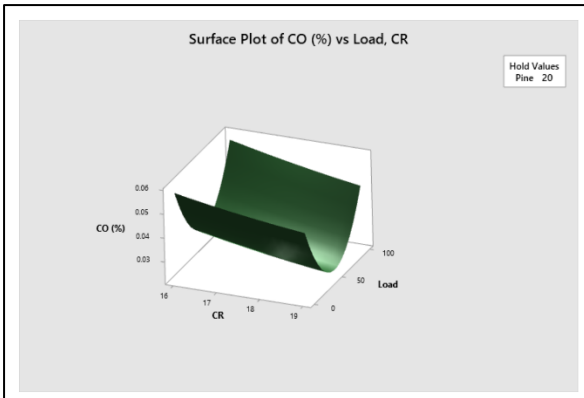


Figure 11 (a)

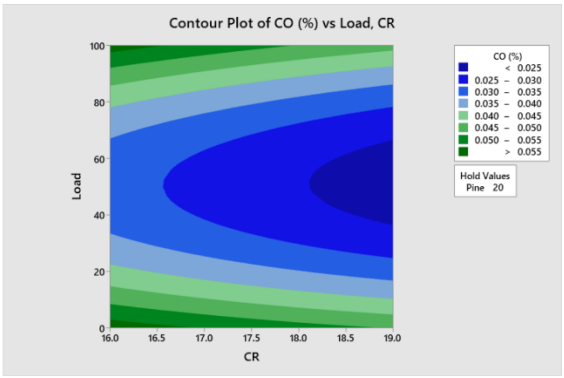


Figure 11 (b)

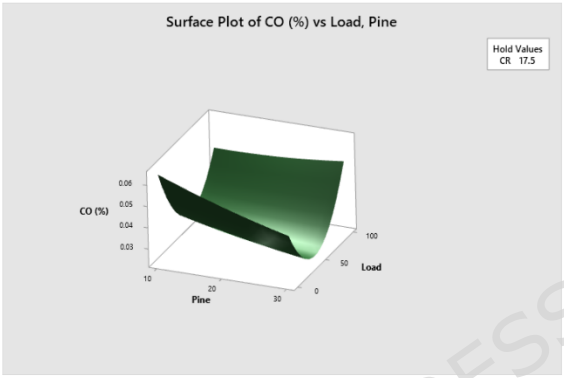


Figure 11 (c)

Figure 11. Response surface and contour plots of carbon monoxide emissions: (a) effect of compression ratio and engine load, (b) contour map of compression ratio and engine load, and (c) effect of PO proportion and engine load.

$$\begin{aligned} CO\ (\%) = {} & 0.194 - 0.0112\ CR - 0.001122\ Load - 0.00096\ Pine \\ & + 0.00026\ CR \times CR \\ & + 0.000010\ Load \times Load + 0.000011\ Pine \times Pine \\ & - 0.000008\ CR \times Load - 0.000025\ CR \times Pine \\ & + 0.000011\ Load \times Pine \end{aligned} \quad (11)$$

The quadratic regression model for CO emissions the RSM-based quadratic regression model for CO emissions is expressed in Eq. (11), which is very accurate. $R^2 = 0.9389$ The three-dimensional response surface of CO emissions versus compression ratio and load is shown in Figure 11(a), and the corresponding contour map in Figure 11(b). It is evident from these plots that a decrease in CO emissions is significant for an increase in compression ratio and load, up till a certain medium-load level. Then it is slightly higher under full load. As can also be seen from the response surface of CO emissions with the PO proportion and load (Figure 11(c)), the larger the fraction of PO, the more significant the decrease in CO emissions, particularly for the medium engine load. The pronounced curvature of the response surfaces in Figures reflects the statistically significant relationships between compression ratio and load.

This implies that the beneficial impact of a higher compression ratio on CO reduction will be maximised at medium engine load. As shown in Figure 11, the experimental data points are strongly correlated with the RSM-predicted surfaces, indicating that the model for CO emission is robust and reliable. Based on the combined experimental and RSM analysis, the lowest CO emissions during PO–assisted PCCI operation are achieved at higher compression ratios, medium engine loads, and increased proportions of PO. Due to better oxidation kinetics, better mixing, and higher combustion temperatures

3.1.5 Combined Experimental and RSM Analysis of NOx Emissions

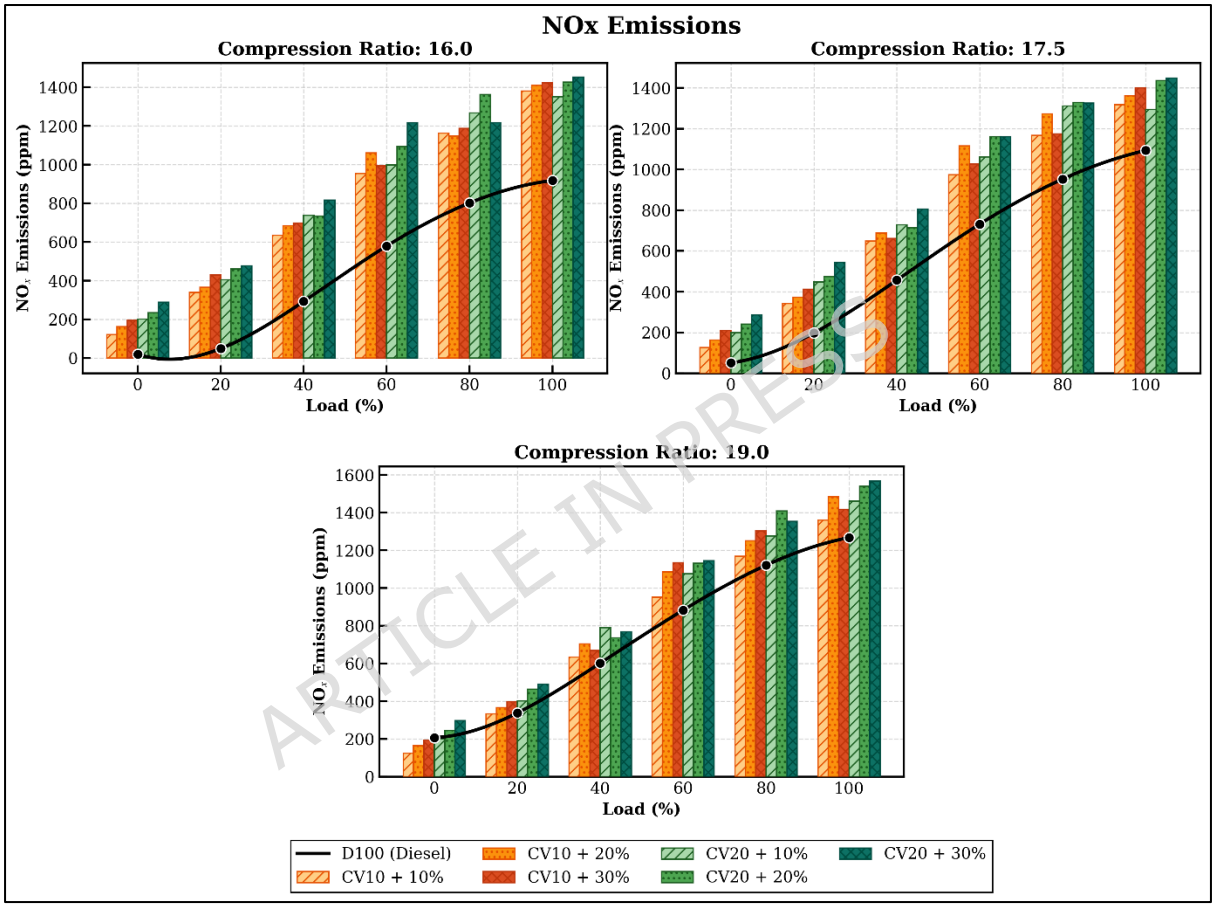


Figure 12. Effect of engine load on NOx emissions under PCCI mode

In the case of D100, as shown in Figure 12, PCCI operation decreased NOx at low to medium engine load, but NOx increased at high engine load irrespective of compression ratio or amount of PO; however, the change was less significant at lower compression ratios, whereas in the presence of other pilot fuels such as CV10 and CV20, NOx emissions increased both at low to medium and high engine loads with lower compression ratios, but a reduction in NOx emissions was seen at higher compression ratios. For all fuel combinations, NOx emissions increase gradually with engine load. The reason is that the rise of the cylinder temperature, the increase of the residence time at high temperatures, and

the Zeldovich mechanism lower the barrier for thermal NOx generation. At very low loads, lean mixtures, low combustion temperatures and very slow oxidation kinetics stop the formation of NOx. The rapid increase of the peak temperature and oxygen availability accelerates the formation of NOx above 40–60% for the load. Using D100 and with a CR of 16, NOx emissions increased from 20 ppm at 0% load to 290 ppm at 40% load, and there was then a sharp further increase to nearly 920 ppm at full load. Now when the compression ratio is increased to CR = 19, the NOx levels increase significantly from approximately 200 ppm at no load to almost 620 ppm at 40% load, peaking at approximately 1270 ppm at 100% load. This indicates that the global NOx increases approximately 35–40% at higher compression ratios due to increased peak cylinder temperature and improved oxygen diffusion. The effect of the addition of PO on NOx emissions as seen from Figure shows NOx emissions increasing from approximately 730 ppm for D100 to about 980 ppm for CV10+P10%, 1,030 ppm for CV10+P20%, and nearly 1,120 ppm for CV10+P30% at CR = 17.5 and 60% load. This represents an increase of approximately 25% to 55% compared to baseline diesel operation. This increment in local flame temperature and subsequently the formation of thermal NOx under PCCI conditions is attributed to the high oxygen content, improved volatility and rapid premixed combustion of PO. CV20 pilot fuel demonstrates a similar trend, where an increase in biodiesel further increases NO due to the higher oxygen content and longer burning period.

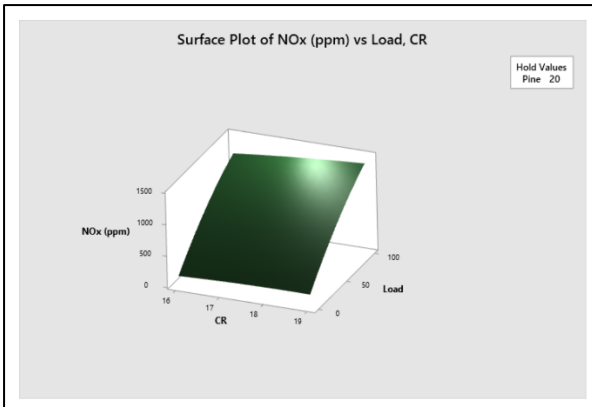


Figure 13 (a)

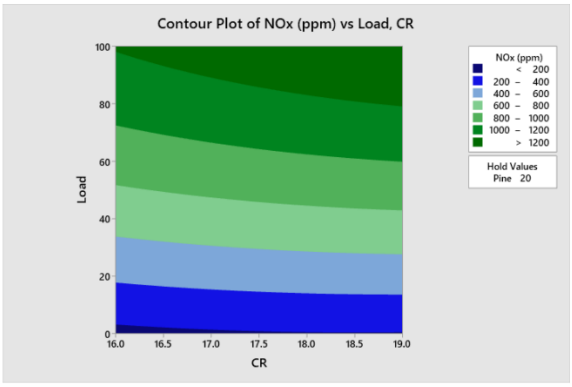


Figure 13 (b)

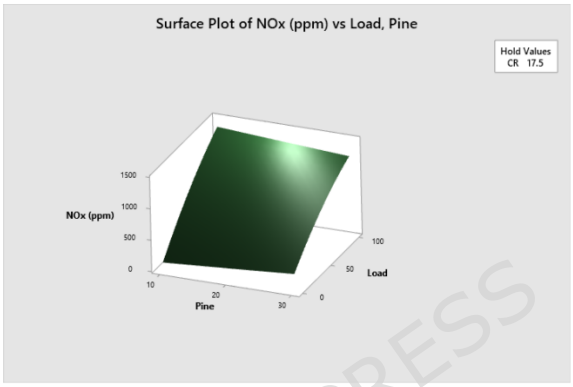


Figure 13 (c)

Figure 13. Response surface and contour plots of NOx emissions: (a) effect of compression ratio and engine load, (b) contour map of compression ratio and engine load, and (c) effect of PO proportion and engine load.

$$\begin{aligned} \text{NOx (ppm)} = & -532 + 140\ \text{CR} + 10.19\ \text{Load} - 69.0\ \text{Pine} - 6.1\ \text{CR}\times\text{CR} \\ & - 0.03930\ \text{Load}\times\text{Load} + 0.020\ \text{Pine}\times\text{Pine} \\ & + 0.448\ \text{CR}\times\text{Load} + 4.325\ \text{CR}\times\text{Pine} - 0.1423\ \text{Load}\times\text{Pine} \end{aligned} \quad (12)$$

Eq. (16), with $R^2 = 0.9967$, indicating an amazing agreement between the predicted values and the experimental values. The three-dimensional response surface of NOx emissions as a function of compression ratio and engine load is depicted in Figure 13(a). The corresponding contour plot is shown in Figure 13(b). These curves indicate very clearly that NOx emissions rise sharply per increase in both compression ratio and load. The maximum slope occurs when the load is greater than 60% and the CR is greater than 18. The NOx emissions response surface to the amount of PO and the load is presented in Figure 13(c). Thus, although the effect of PO on NOx formation is not significant with no load, it is greatly accelerated with increasing amounts of PO with a full load on the engine. The statistically significant nature of the combined interaction between compression ratio and load can be clearly seen from the high curvature response surfaces. Thus, it suggests that the synergistic effect of a higher compression ratio and the higher load increase the thermal NO

formation[44]. The near congruence between the experimental data points and the RSM-predicted surfaces indicates that the NOx emission model is robust and reliable. Experimental and RSM results together reflect maximum NOx emissions occurring at higher compression ratios, engine loads and PO proportions during PCCI operation of PO. This is primarily due to increased combustion temperature, the use of oxygen and improved thermal kinetics of the NOx reaction.

3.1.6 Combined Experimental and RSM Analysis of Smoke Opacity

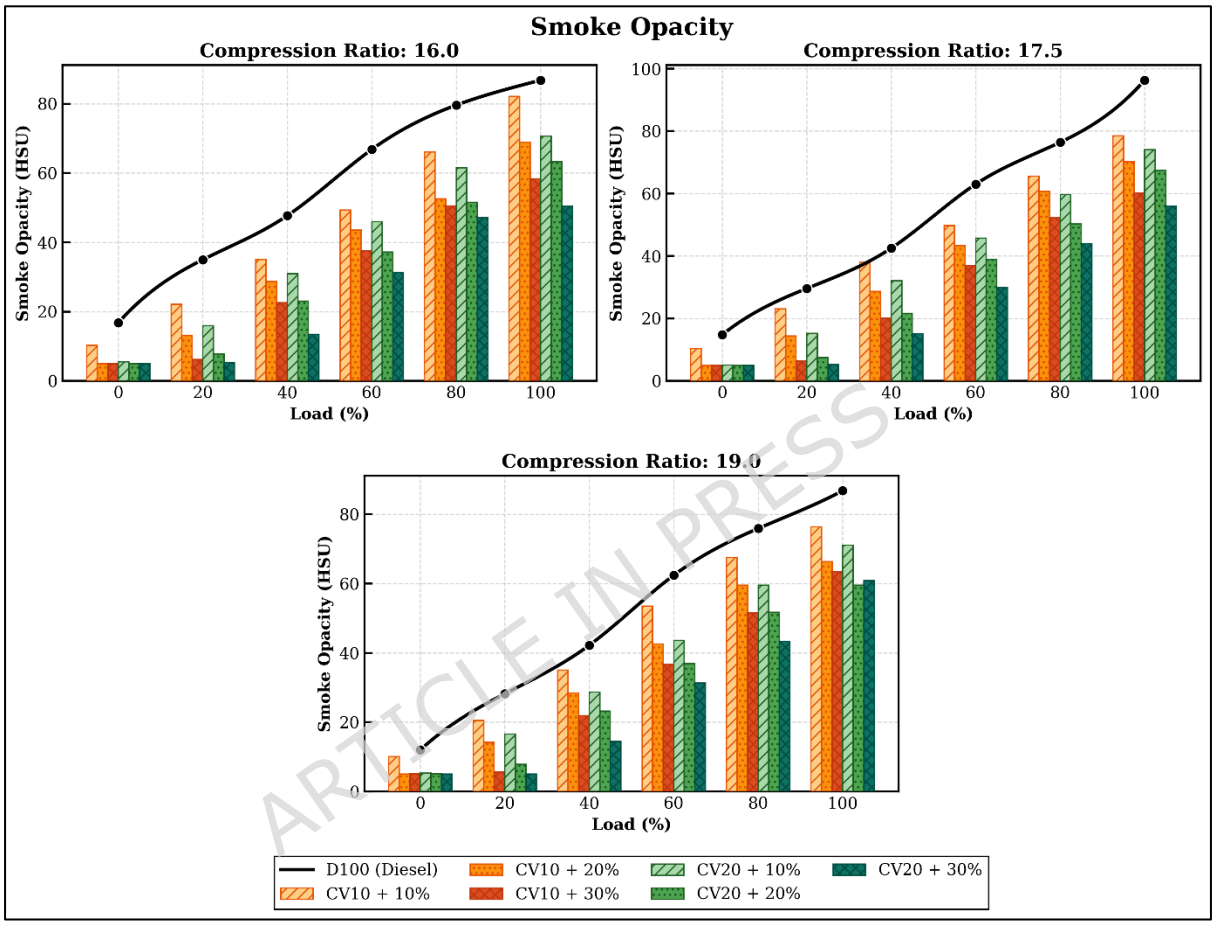


Figure 14. Effect of engine load on Smoke Opacity under PCCI mode

Figure 14 presents the variation of smoke opacities with engine load for different combinations of compression ratios and amounts of PO when PCCI has been applied using D100, CV10 and CV20 as pilot fuels. All operating points show a steady increase in smoke opacity with an increase in engine load. At larger loads more fuel is injected; locally rich regions are created with diffusion-controlling combustion at high load levels. At low loads, the smoke levels remain low because the mixtures are lean, there is plenty of oxygen, and the soot precursors are fully oxidised under PCCI conditions. At loads greater than 60%, both pressure and air-fuel equivalence ratio, as well as the tendency to promote soot nucleation, soot surface growth and agglomeration, increase in the cylinder. It ends up

releasing more smoke. For CR 16 and clean diesel (D100), smoke opacity increases from approximately 16.8 HSU at 0% load to about 47.7 HSU at 40% load, then increases rapidly to nearly 86.8 HSU at full load. At the highest compression (CR = 19), all loads had more transparent smoke. It goes from 12.0 from no load through 62.4 at 60% load and all the way up to ~86.8 at full load. Increased compression ratios result in higher temperatures and increased mixing within the cylinder, which addresses the high soot yield and soot particle combustion at the expansion stroke, leading to an effective reduction of smoke by about 20–25% at intermediate loads. It illustrates dramatically in Figure how the addition of PO affected the volume of smoke emitted. When the load is 80% and the CR is 17.5, the smoke opacity decreases from approximately 76.4 HSU for D100 to approximately 65.6 HSU for CV10+P10%, 60.7 HSU for CV10+P20% and finally to a minimum value of approximately 52.2 HSU for CV10+P30%. This represents a reduction of approximately 32% in smoke opacity compared to baseline diesel operation. This significant improvement is attributed to the high volatility, low viscosity, and fuel-bound oxygen content of PO, which collectively enhance fuel atomization, vaporization, and the subsequent soot oxidation rate during premixed combustion. At high pine oil replacement (P30) and maximum load of the engine, slight smoke was observed due to the locally rich combustion resulting in insufficient time for soot to oxidise completely. CV10 (the lowest smoke opacity of the pilot fuels) is always at the top of the column, followed by D100 and CV20. At CR = 19 and 80% load, smoke opacity values are approximately 75.9 HSU for D100, 67.5 HSU for CV10+P10% and 59.5 HSU for CV20+P10%. Instead, CV10 balances oxygenation and fuel breakdown better. Because of the increased viscosity and density of higher biodiesel blends, the higher smoke levels were observed with CV20. We believe these blends increase the difficulty for the spray to atomize, increase the amount of fuel making contact with the cylinder walls, and facilitate increased soot precursor formation during the high-load PCCI operation of the engine.

$$\begin{aligned}\text{Smoke (HSU)} = {} & -114 + 12.1\,CR + 1.052\,Load + 1.75\,Pine - 0.266\,CR\times CR \\ & + 0.002211\,Load\times Load + 0.0003\,Pine\times Pine \\ & - 0.0317\,CR\times Load - 0.1200\,CR\times Pine \\ & - 0.00220\,Load\times Pine \end{aligned} \quad (13)$$

The developed RSM-based quadratic regression model for smoke opacity (Eq. (13)) has a coefficient of determination greater than $R^2 = 0.9953$, confirming the strength of the regression between experimental and predicted values. Smoke opacity (SO) three-dimensional response surface Figure 15 (a) as a function of compression ratio and engine load The corresponding contour plot is depicted in Figure 15(b). It is pretty evident from these plots that the smoke opacity decreases with an increase in the

compression ratio and increases considerably with an increase in the load. The response surface of smoke opacity as a function of PO and load is illustrated in Figure 15(c). This indicates that as the quantity of PO increases, smoke formation drops dramatically until moderate levels of loading are reached. Then, the part of decrease slows down due to the fact that combustion is diffusion-controlled. Statistical significance tests show that the interaction between compression ratio and load is significant, which is evident from the curvature of the response surfaces. It implies that the positive influence of a higher compression ratio in reducing smoke is most pronounced at the medium load (40–60%) range. The smoke model is considered to be powerful and reliable for predictions due to the high degree of correlation between the experimental data and the RSM-predicted response surfaces depicted in Figs. 15(a-c). Through the combination of experimental results and RSM analysis, the most favourable smoke emissions during PO–assisted PCCI operation are found at high compression ratios, at medium engine loads and with high PO levels due to better premixing, better fuel oxidation and less soot precursor formation, respectively.

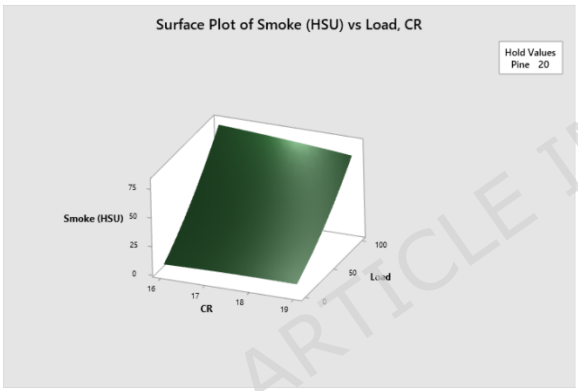


Figure 15 (a)

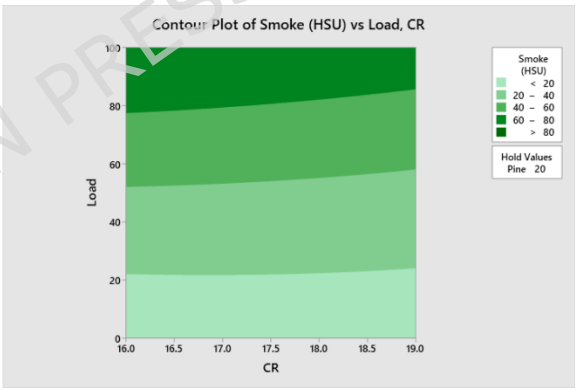


Figure 15 (b)

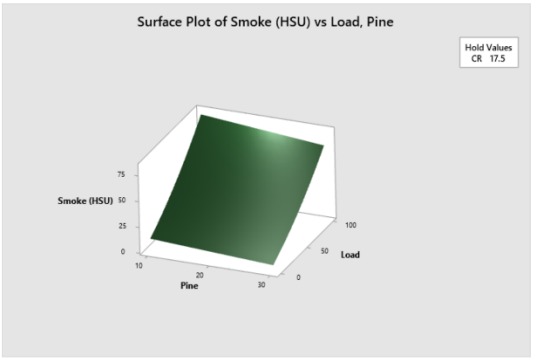


Figure 15 (c)

Figure 15. Response surface and contour plots of NOx emissions: (a) effect of compression ratio and engine load, (b) contour map

of compression ratio and engine load, and (c) effect of PO proportion and engine load.

Based on the individual analyses of performance and emission characteristics discussed above, a consolidated comparative assessment was carried out to clearly isolate the effects of compression ratio and pine oil substitution under PCCI operation. The following figures 16 a and 16 b summarize these isolated effects using representative operating conditions.

The comparative plots clearly demonstrate that compression ratio predominantly governs thermal efficiency and NO_x formation, whereas pine oil substitution has a stronger influence on smoke opacity and unburned hydrocarbon emissions. These experimental trends provide the basis for subsequent response surface modelling, machine learning prediction, and quality-based optimization presented in the following sections.

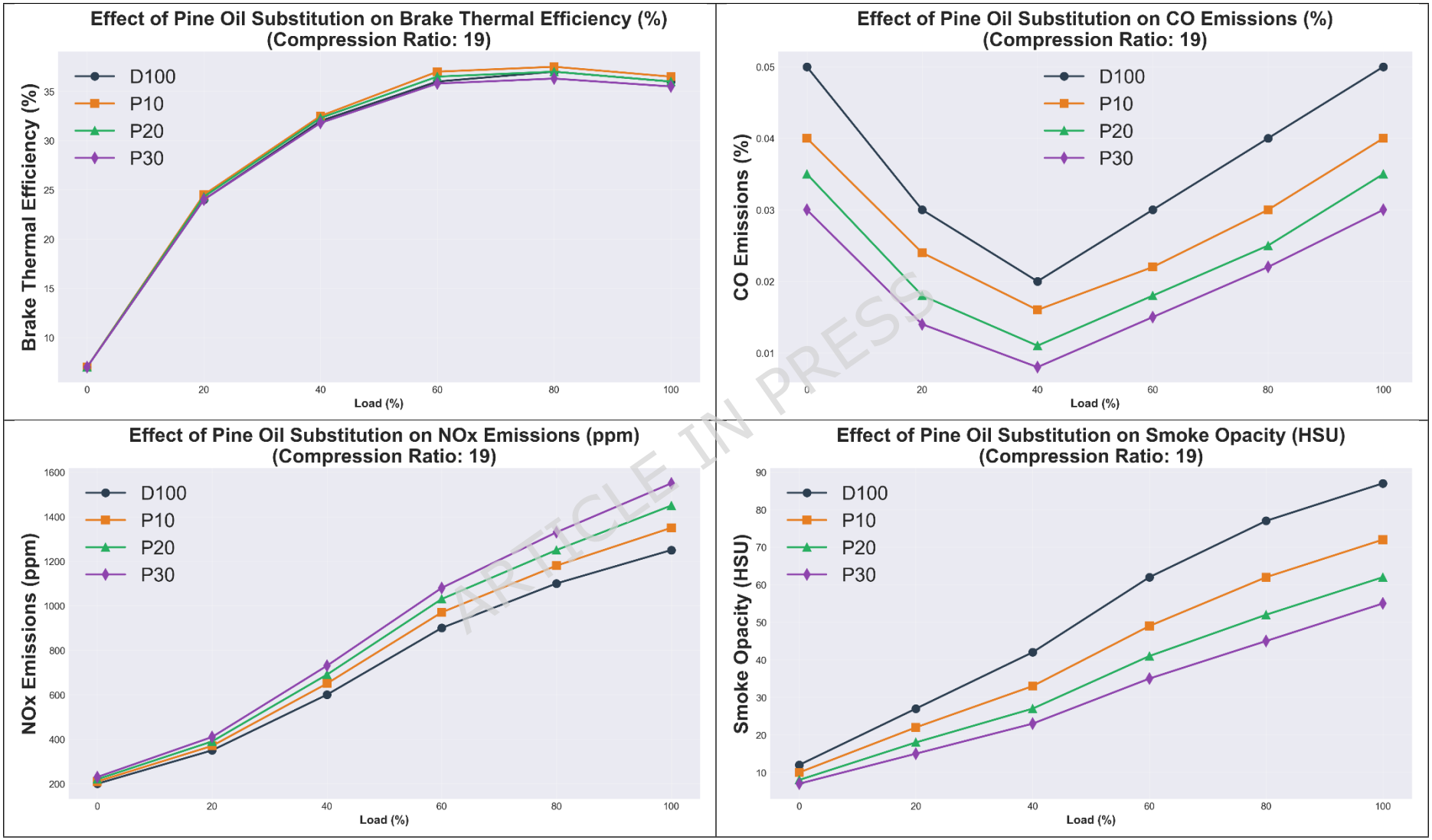


Figure 16a. Effect of pine oil substitution on PCCI engine performance and emissions at CR = 19

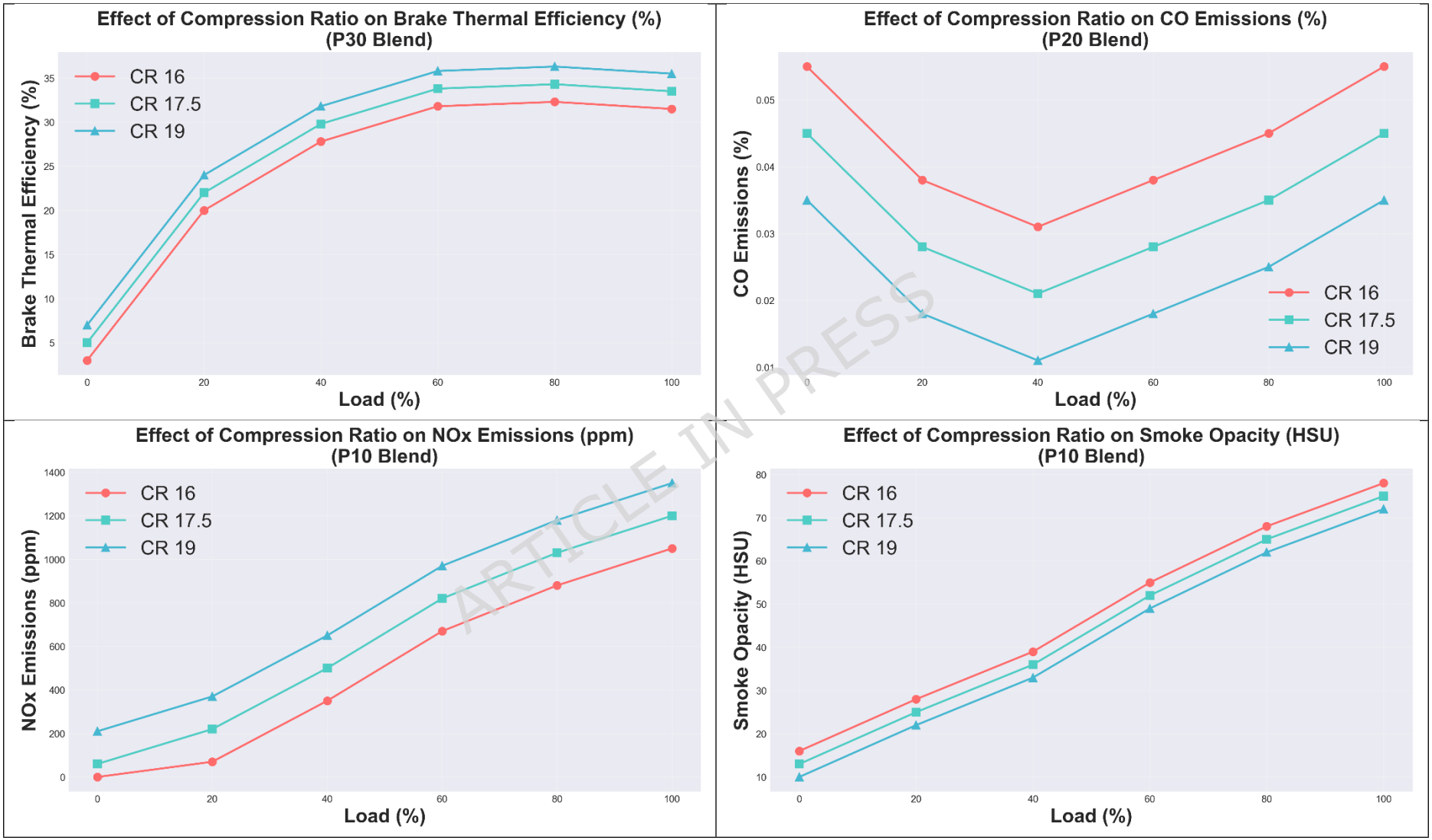


Figure 16 b. Effect of compression ratio on PCCI engine performance and emissions for CV10 + PO30 blend.

3.1.7 RSM Model Development, Adequacy Assessment, and ANOVA Validation

The results shown in Table 5 show that developed quadratic response surface methodology models performed satisfactorily in terms of brake thermal efficiency, brake specific fuel consumption, hydrocarbon emissions, carbon monoxide, nitrogen oxides, and smoke opacity. Even though the coefficients of determination (R^2) were high, the assessment of model adequacy did not rely solely on R^2 values. The anticipated R^2 values closely aligned with the adjusted R^2 values across all responses. The findings indicate that the model successfully generated precise predictions while avoiding overfitting. Furthermore, the results of the lack-of-fit tests were statistically non-significant ($p > 0.05$) across all models, thereby validating that the quadratic formulations effectively capture the experimental data within the defined design space. The analysis of residual diagnostics, such as normal probability plots and the relationship between residuals and fitted values, revealed no systematic trends, thus confirming the model's robustness.

All regression models had p-values below 0.0001, demonstrating a high degree of statistical significance. The difference between the adjusted R^2 and predicted R^2 values was always less than 0.2 for all of the responses. The results suggest that the RSM model is reliable enough to be applied. The contribution of experimental error was always less than 3% for all responses, highlighting good repeatability and measurement consistency. The variance analysis indicated that the influences of linear load, pine oil fraction, and quadratic load significantly affected HC emissions. Conversely, it was determined that quadratic load terms are the primary contributors to CO emissions, highlighting the significance of engine load in PCCI operation. NO_x was primarily influenced by effects from linear loads but also significant interaction effects between compression ratio and pine oil fraction. These effects highlight the role that temperature-dependent kinetics and fuel-bound oxygen interactions can play. The load on the engine was the main parameter affecting the density of the smoke. The influence of the pine oil fraction and the interplay between compression ratio and load appeared to be limited, suggesting that higher compression ratios promote soot oxidation.

Table 5. RSM Model Adequacy Statistics for All Responses

Response	R^2	Adj-R^2	Pred-R^2	Model p-value	Lack-of-Fit
BTE	0.9999	0.9989	0.9989	<0.0001	Not significant
BSFC	0.9999	0.9998	0.9997	<0.0001	Not significant

HC	0.9717	0.9631	0.9524	<0.0001	Not significant
CO	0.9743	0.9641	0.9578	<0.0001	Not significant
NOx	0.9967	0.9948	0.9926	<0.0001	Not significant
Smoke	0.9953	0.9928	0.9904	<0.0001	Not significant

The results indicate that the model provides accurate predictions without evidence of overfitting. Although very high coefficients of determination were obtained, it is emphasized that the validity of the developed RSM models is confined to the investigated operating domain defined by the tested compression ratios, engine loads, and pine oil substitution levels. Extrapolation beyond this experimental space should therefore be undertaken with caution. The optimised operating condition derived from the multi-response desirability method (CR = 18.94, load = 43.43%) was experimentally validated at the closest discrete operating point (CR = 19, load = 40%). The experimental results presented in Table 6 closely align with the predictions made by RSM. The majority of the variations fall within the 2.1% range of experimental uncertainty. The variations in BSFC and smoke opacity can be elucidated by the sensitivity of premixed PCCI combustion to the addition of increased pine oil and minor alterations in the fuel mixing process within the cylinder. The findings validate the accuracy and real-world applicability of the developed RSM models.

Table 6. RSM Model Adequacy Statistics for All Responses

Response Parameter	Optimised Value (RSM Prediction)	Experimental Value	Absolute Deviation	Deviation (%)
BTE (%)	30.787	**31.47**	0.68	**2.21**
BSFC (kg/kWh)	0.3635	**0.29**	0.0735	**20.2**
CO (%)	0.01947	**0.017**	0.00247	**12.7**
HC (ppm)	31.57	**26**	5.57	**17.6**
NO_x (ppm)	881.2	**768**	113.2	**12.8**
Smoke (HSU)	24.74	**14.4**	10.34	**41.8**

3.2 – Machine Learning Results and Validation

3.2.1 Machine Learning Metrics Evaluation

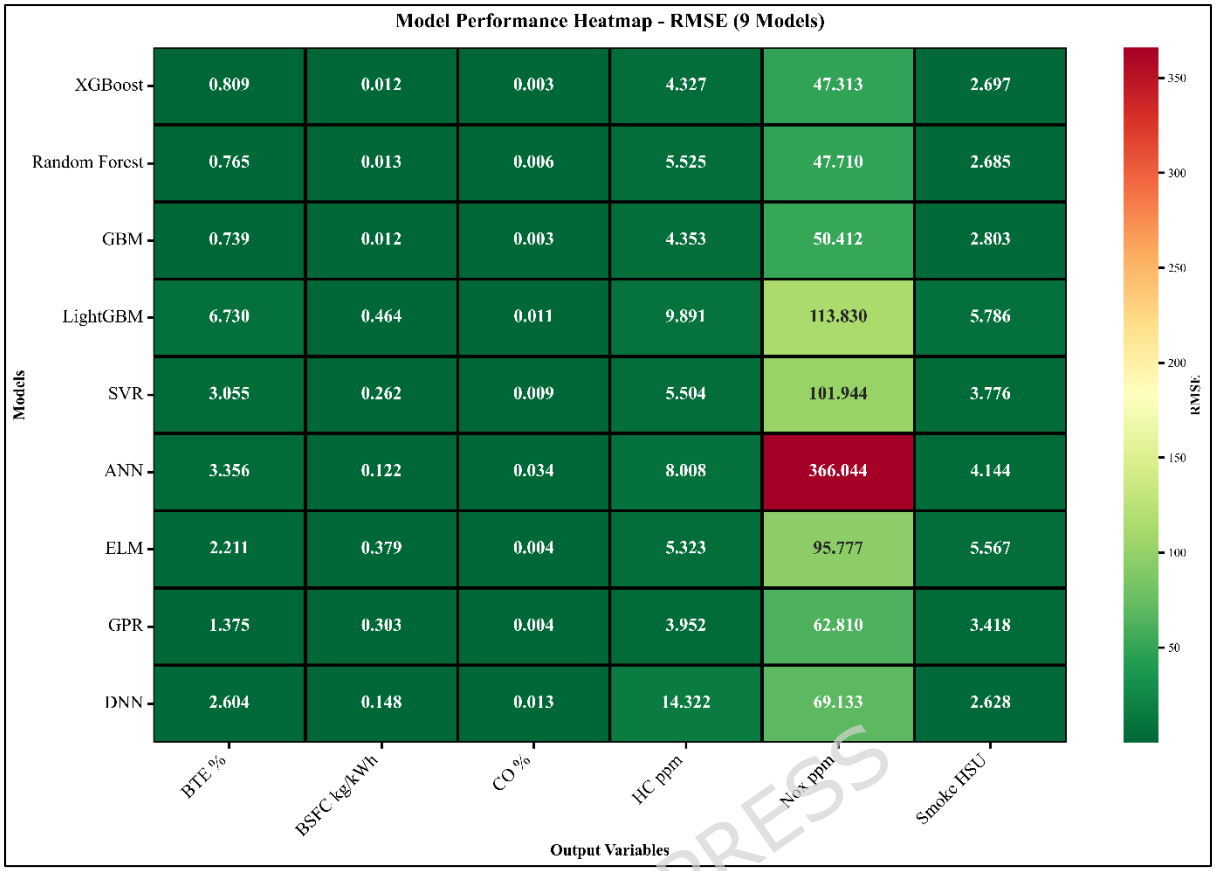


Figure 17: Model performance heatmap – RMSE

First, the nine machine learning models were compared in terms of performance using seven standard statistical metrics, such as R^2, root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), Pearson's correlation coefficient (r), mean bias error (MBE) and normalised RMSE (NRMSE). We have grouped the RMSE across all output variables, and we can see the distribution in Figure 17 (Model performance heatmap – RMSE), and likewise with the R^2 in Figure 19 (Model performance heatmap – R^2). Figure 18 (Overall model performance comparison) shows a consolidated bar-wise comparison of R^2, RMSE, MAE, MSE, MAPE, Pearson-r, MBE, and NRMSE for all nine models. Table 6 (Overall ML metric comparison for all outputs) summarises the exact numerical values used for these comparisons.

Table 7. Overall machine learning performance metrics for all output variables

Model	BTE – RMSE	BSFC – RMSE	CO – RMSE	HC – RMSE	NOx – RMSE	Smoke – RMSE	BTE – R^2	BSFC – R^2	CO – R^2	HC – R^2	NOx – R^2	Smoke – R^2
XGBoost	0.809	0.012	0.003	4.327	47.313	2.697	0.994	0.999	0.954	0.927	0.987	0.989
Random Forest	0.765	0.013	0.006	5.525	47.71	2.685	0.995	0.999	0.846	0.881	0.987	0.989
GBM	0.739	0.012	0.003	4.353	50.412	2.803	0.995	0.999	0.959	0.926	0.985	0.988
LightGBM	6.73	0.464	0.011	9.891	113.83	5.786	0.586	0.066	0.389	0.617	0.923	0.95
SVR	3.055	0.262	0.009	5.504	101.944	3.776	0.915	0.701	0.626	0.881	0.938	0.979
ANN	3.356	0.122	0.034	8.008	366.044	4.144	0.897	0.935	-4.843	0.749	0.206	0.975
ELM	2.211	0.379	0.004	5.323	95.777	5.567	0.955	0.375	0.922	0.889	0.946	0.954
GPR	1.375	0.303	0.004	3.952	62.81	3.418	0.983	0.602	0.936	0.939	0.977	0.983
DNN	2.604	0.148	0.013	14.322	69.133	2.628	0.938	0.905	0.174	0.198	0.972	0.99

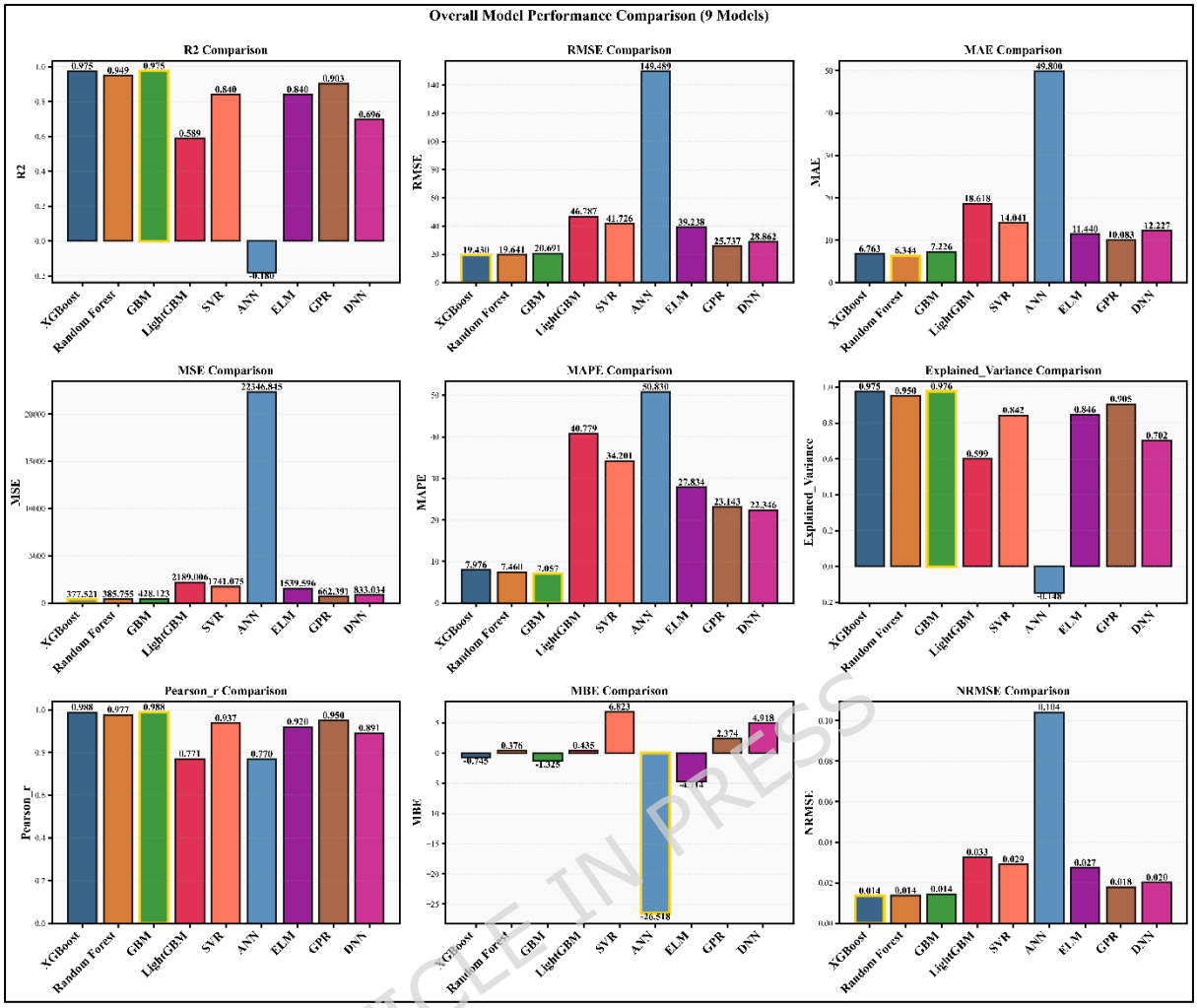


Figure 18: Overall model performance comparison using R^2, RMSE, MAE, MSE, MAPE, Pearson-r, MBE, and NRMSE

The highest R^2 values for most performance and emission outputs are always observed in the Gradient Boosting Machine (GBM), XGBoost and Random Forest (RF) models. It means they can learn and generalise more effectively. For BTE Lin GBM is the most accurate model with an R^2 = 0.995, RMSE = 0.7385, and MAE = 0.6345. The result indicates that it has high predictive power for efficiency. For BSFC, XGBoost was also the best model, producing R^2 and RMSE values equal to 0.9994 and 0.0116 kg/kWh, respectively.

The performance from these models was fairly consistent. Among the algorithms, GBM did the best for CO (R^2 = 0.9588, RMSE = 0.0029%) and was highly sensitive to changes in combustion quality. All response plots shown in this paper suggest that the best-performing model for predicting HC emissions was the Gaussian Process Regression (GPR) model. This can deal with nonlinear emission behaviour very well, as can be seen from the R^2 of 0.9389 and RMSE of 3.95 ppm. For nitrogen oxides (NOx), XGBoost again exhibited greater accuracy over the other

models, in terms of R^2 = 0.9867 and RMSE of 47.31 ppm. Among the three models, the Deep Neural Network (DNN) performs best with R^2 = 0.9898 and RMSE = 2.63 HSU in terms of smoke opacity measurement.

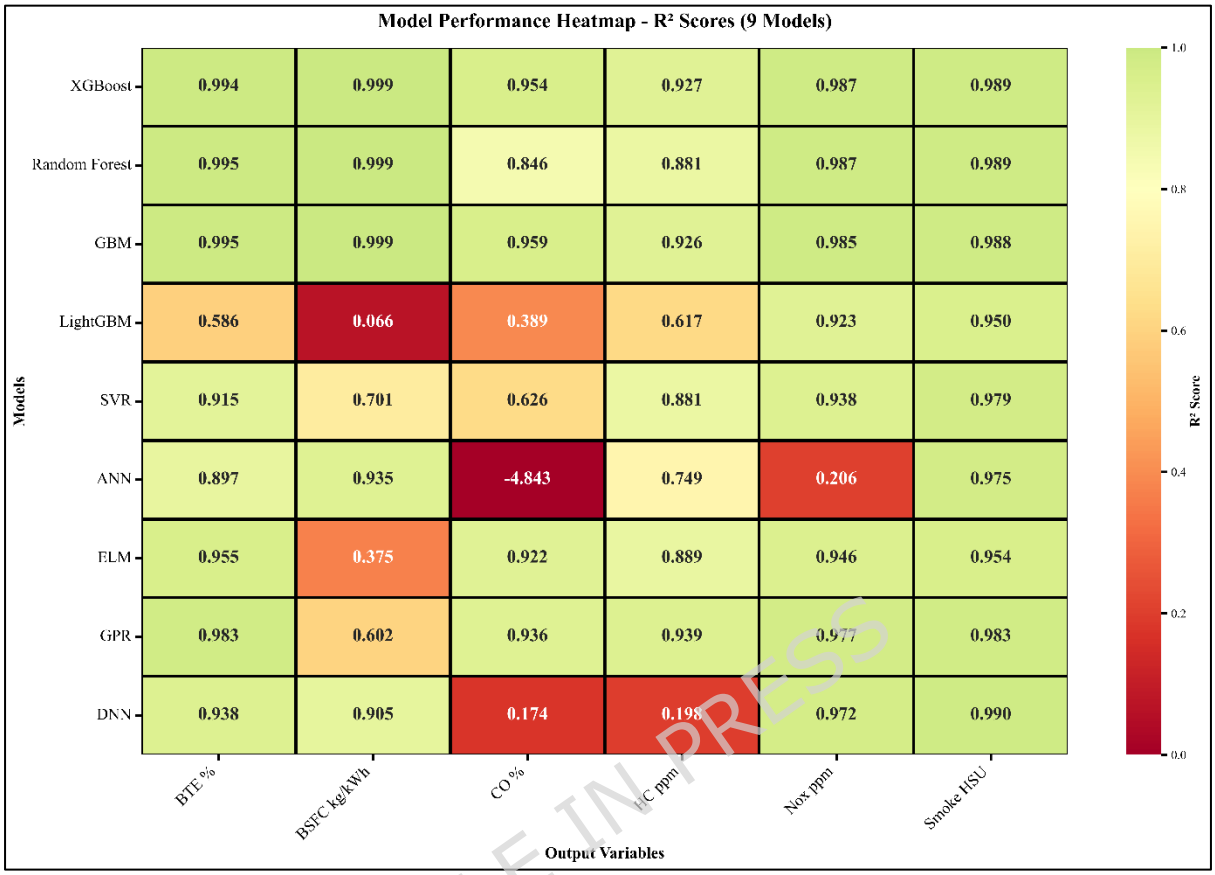


Figure 19: Model performance heatmap – R^2 scores

The comparative R^2 trends shown in Figure 18 indicate that ANN and LightGBM yielded higher values for RMSE, MAE, and MAPE in some outputs, particularly for RMSE, MAE, and MAPE with respect to NOx and HC. That means they do not generalise quite so well. The ANN model shows no stability for small experimental datasets with negative or very low R^2 values for CO and NOx predictions. For the best-performing models, Pearson's correlation coefficients are above 0.98, indicating a strong linear relationship between the predicted and experimental values. The low NRMSE (with < 0.03 for most of the best-performing models) validates the ML framework as statistically not providing large errors.

On the bases of this metric-based evaluation, we can see that the best predictive models for current PCCI engine dataset are GBM, XGBoost and Random Forest. In contrast, GPR and DNN are tailored for high-fidelity HC and smoke emissions prediction, respectively. The extensive statistical filtering provides deliberate reasons with numbers to best pick the best of the ML models to use in subsequent snorkelling and validation/optimisation analysis. Therefore, judging from the completely statistical data in Table 7, Figure 17, Figure 18, and Figure 19, it can be said that GBM is the most robust and consistent model for simultaneous

prediction of performance and emissions with the assistance of PO for PCCI operation. There's a slight difference that put Random Forest and XGBoost more close to each other.

3.2.2 Predicted vs. Actual Validation of Brake Thermal Efficiency (BTE)

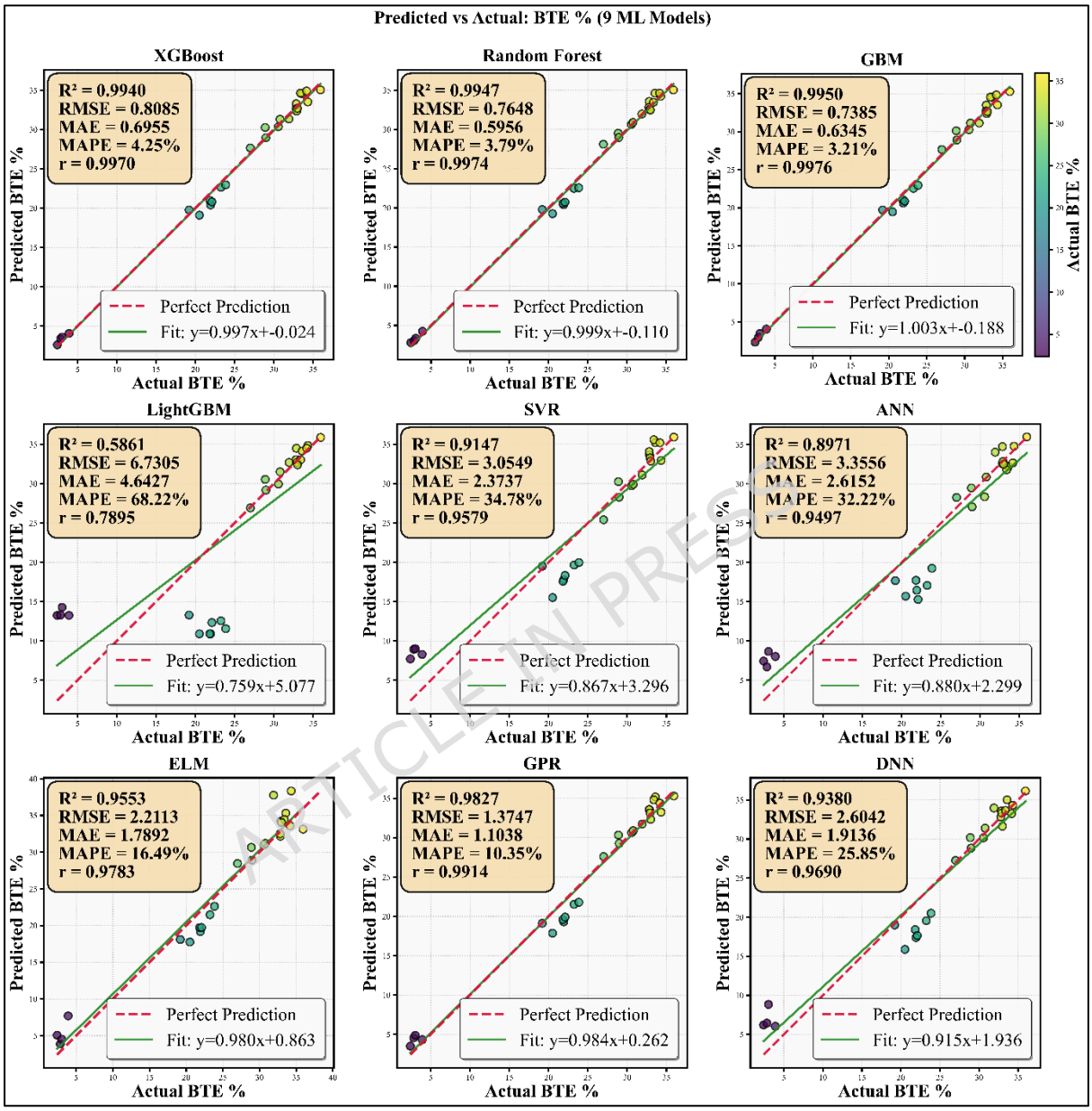


Figure 20. Predicted vs. Actual Validation of Brake Thermal Efficiency (BTE)

The predicted vs actual comparison for BTE using nine different machine learning models is depicted in Figure 20. The comparison demonstrates that the ensemble-based models, including XGBoost, Random Forest and Gradient Boosting Machine (GBM), yield extremely high predictive accuracy, $R^2 > 0.994$, and very low values of root mean square error (RMSE) and mean absolute error (MAE). Over the full operating range, the predicted BTE values for these models sit very near

the 45° reference line. Which implies little systematic bias and well-generalising models. Tree-based models have a great predictive capability, as they can closely approximate the nonlinear correspondence between input variables (load, compression ratio and fuel blend ratio, etc.) and output level (thermal efficiency). Moreover, they also excel at capturing feature interactions, which are essential for a combustion system where the phenomenology is governed by complex thermodynamics and complicated chemical kinetics. Similarly, the Gaussian Process regression (GPR) and Extreme Learning Machine (ELM) models correlate well with the experimental values in general conduct but have characteristics in the middle of the load range leading to small differences.

However, we can observe much clearer scattering at higher BTEs, especially around peak efficiency zones, compared with LightGBM and ANN. The R^2 is lower and the RMSE is higher. This clearly shows that those models drop their reliability at combustion values that are very high (i.e., small variations in the experiments can lead to very large variation in efficiency). Because the SVR is sensitive to kernel parametric choices, it can handle nonlinearity but tends to underpredict at extreme BTE. GBM yields the best-performing model overall for BTE prediction, as shown in the predicted vs actual analysis. It has the best tradeoff between accuracy, stability and compute speed. The BTE values predicted using the developed ML framework and those measured experimentally are close to each other, confirming that the developed ML framework can predict power efficiency in a wide range of cyclic PCCI operating conditions. . Similar to the RSM analysis, the predictive capability of the machine learning models is limited to the operating conditions represented in the experimental dataset.

3.2.3 Predicted vs. Actual Validation of Brake Specific Fuel Consumption (BSFC)

BSFC is one of the driving factors in developing an internal combustion engine with optimal performance. Figure 21 demonstrates the results of all nine machine learning models for BSFC in terms of predicted and experimental values. Out of all the models, the best-known models to use for better prediction are XGBoost, Random Forest, and GBM. The R^2 values approach 0.999, and the RMSE values are well below 0.013 kg/kWh. The data points for all of these models nearly fall on the 1:1 reference line, demonstrating that all models agree very well across the entire range of BSFC. This works so well because of the ensemble learning structure. The way that tercile reflects the nonlinear nature of BSFC dependency on load, fuel composition and combustion phasing is captured quickly. In addition to it, the low values of MAPE and MAE indicate the relatively low percentage of error, which makes these models also very

important to predict and optimise fuel economy. Although the GPR and DNN models are reasonably predictive, differences remain in terms of better performance at high BSFCs under low load and rich mixture.

The LightGBM, SVR, and ELM models, however, have a larger spread, particularly for the high BSFC case. This is because it is borderline impossible to precisely map angular nonlinear shifts in fuel consumption on very high outputs. ELM, moreover, has a transparent error in its prediction due to single-step learning, which might not be sufficient to learn some interactions with the fuel and combustion as a whole. As expected, the predicted vs total BSFC analysis provides an even more corroborative result concerning the statistical results. It also indicates that XGBoost and GBM are the optimum models for prediction of fuel consumption when applying PCCI. The validation results demonstrate the applicability of the developed ML framework for online estimation of BSFC and potential fuel economy improvements in advanced CI engines.

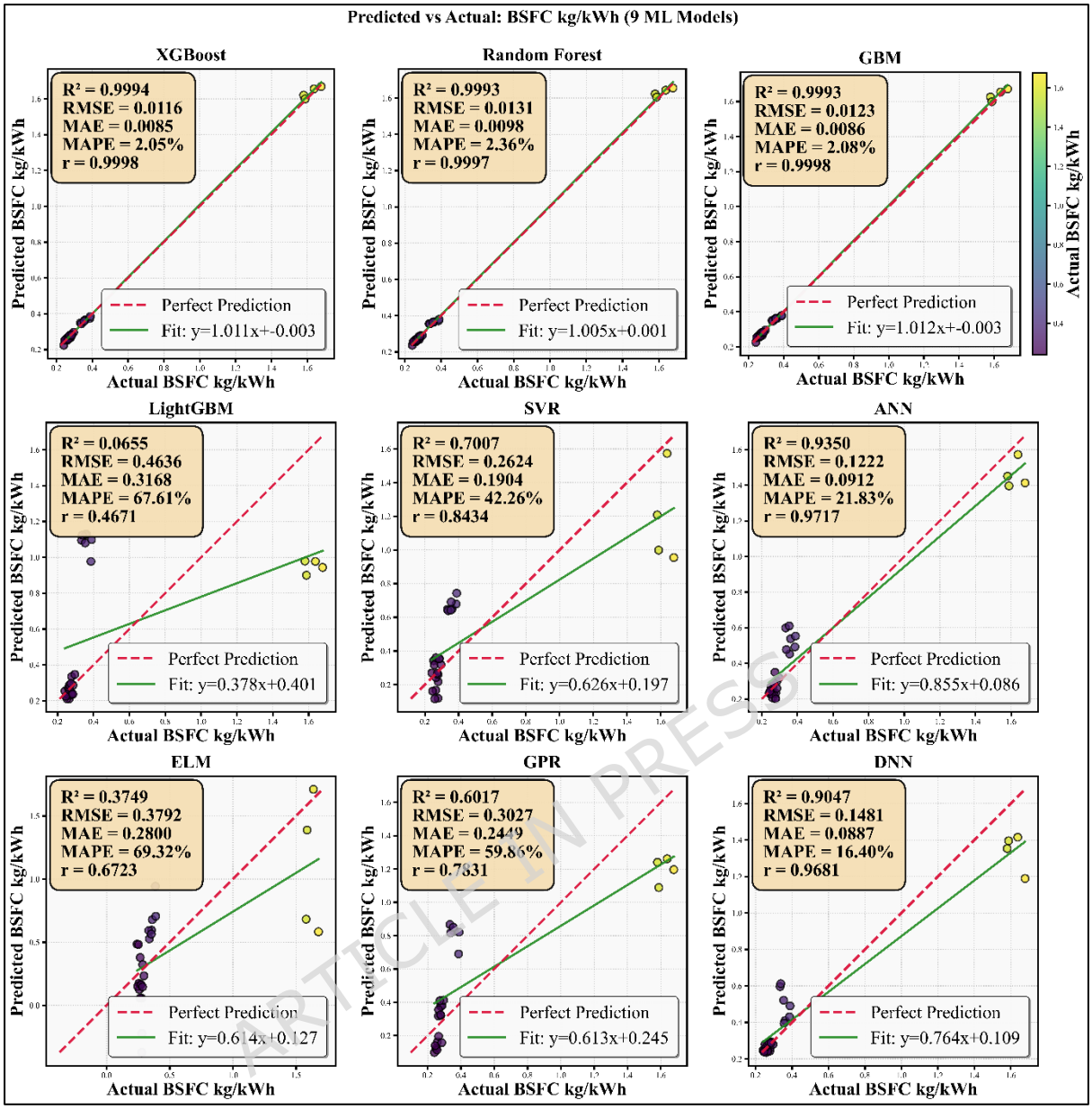


Figure 21. Predicted vs. Actual Validation of Brake Specific Fuel Consumption (BSFC)

3.2.4 Predicted vs. Actual Validation of Unburned Hydrocarbons (HC)

All nine ML models showed a good fit in the predicted vs. actual HC emission plots (Figure 22). Results reveal that R^2 values higher than 0.92 and a small RMSE across the entire operating range for these three models prove the predictive accuracy of XGBoost, Random Forest and GBM methods. Indeed, the estimated HC values from these models closely follow the 1:1 reference line, indicating that they are good at modelling incomplete oxidation processes. HC formation in PCCI engines is severely dependent on mixture quality, wall quenching, ignition delay and low-temperature regions. This quality of ensemble models enables it to predict these nonlinear combustion irregularities quite accurately, hence

excellent matching with experimental data. For the mid-emission range, the SVR and GPR models perform fairly well too, but they show a mild underprediction under higher HC levels occurring at low loads with high EGR.

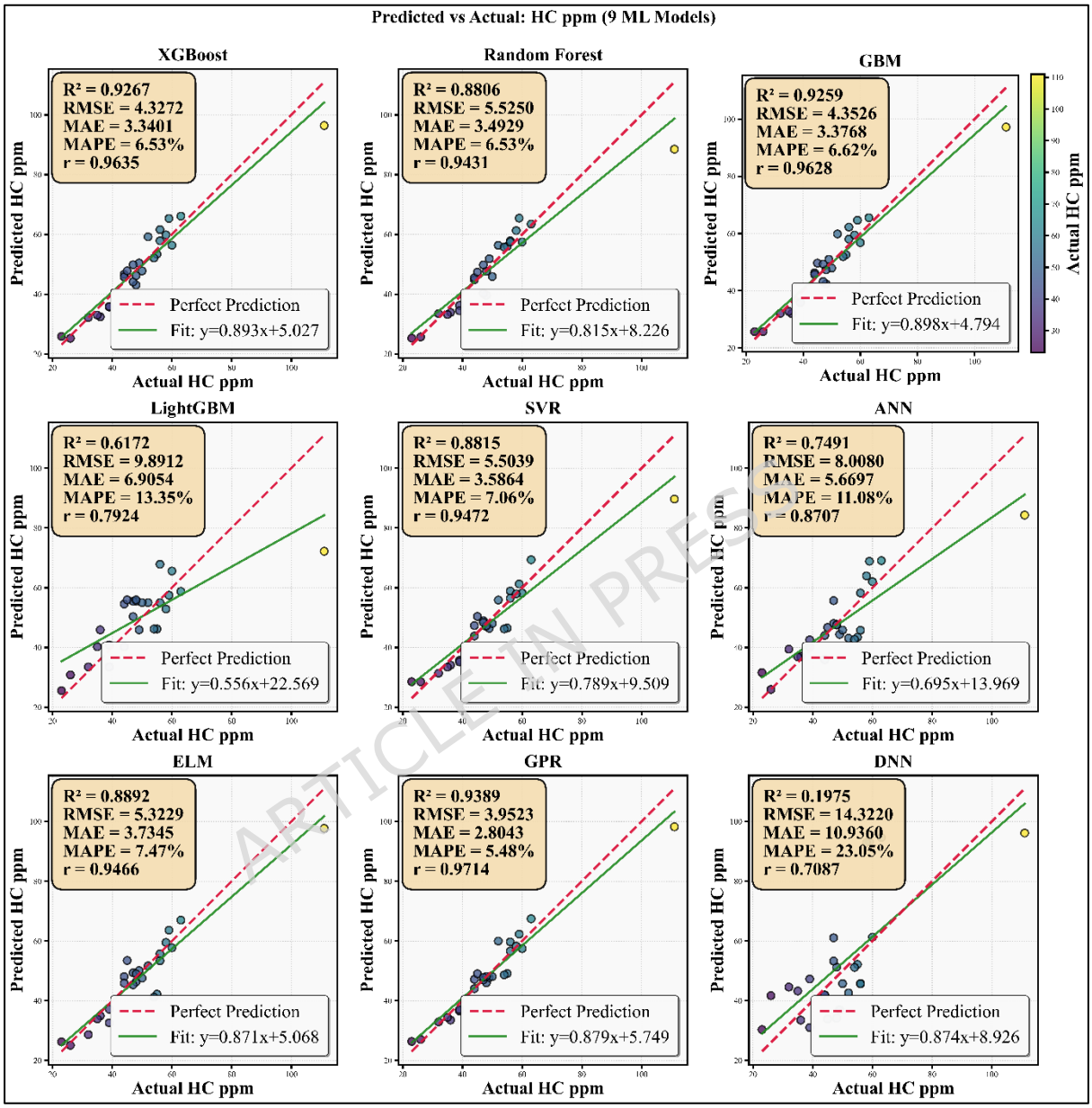


Figure 22. Predicted vs. Actual Validation of Unburned Hydrocarbons (HC)

The scatter among the LightGBM, ANN and DNN models is quite high, particularly in lower to medium HC. Indicates that these models necessarily do not perform well where a regime with frequent misfires is continuous with a regime of complete combustion. This leads to large classification errors, but the LMB model is still fast to compute because of its highly parallelized learning structure. The parity analysis directly supports that XGBoost provides the most uniform and accurate prediction

of HC emission across all engine loads. Hence, it justifies that it is good to optimise PCCI modes of engines sensitive to emissions with blends of oxygenated fuel.

3.2.5 Predicted vs. Actual Validation of Carbon Monoxide (CO) Emissions

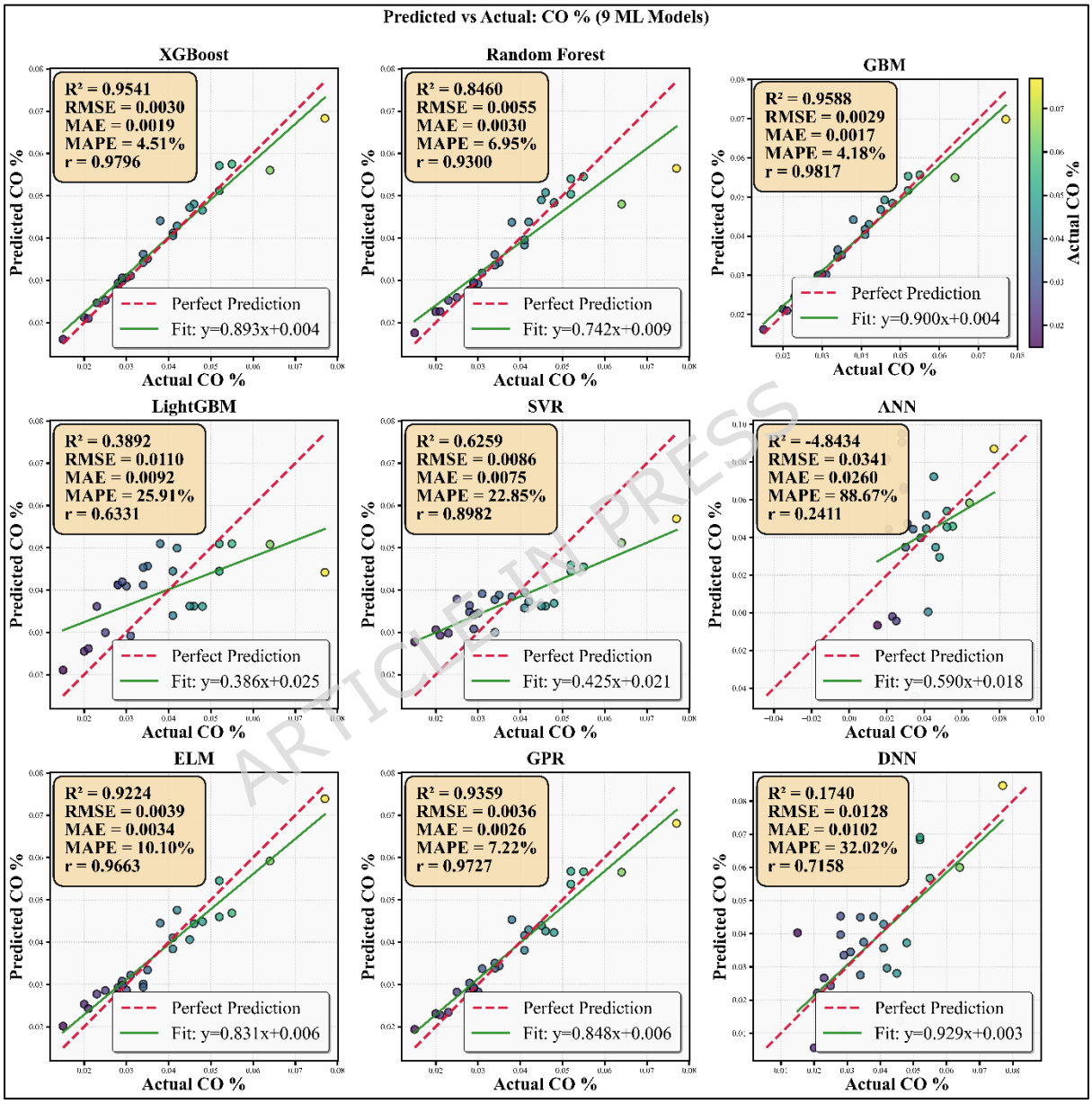


Figure 23. Predicted vs. Actual Validation of Carbon Monoxide (CO) Emissions

Figure 23 display the parity plots and error statistics for HC emissions for the ML models, respectively. The findings indicate that XGBoost, GBM, and Random Forest perform an excellent prediction of CO emissions. The data points are very close to the ideal prediction line (the dotted line), and the R^2 values are more than 0.95. Their lower RMSE and

MAE values reflect how they would be able to estimate the incomplete combustion accurately. Output of CO from a car mostly depends on the local equivalence ratio, fuel atomisation, availability of oxygen and combustion temperature. Ensemble models outperform other types of models due to this ability to capture the nonlinear interaction effects of multiple variables. The GPR and ELM models also performed reasonably well, although these models overpredicted at low and moderate CO levels, and underpredicted at higher levels.

In contrast to SVR, LightGBM and ANN appear scattered, particularly in the low-CO region, where small mixtures and ignition delays lead to a high level of sensitivity in combustion. More importantly, ANN predictions are not very stable, due to the sensitivity of small training data and also due to overfitting. The DNN model is able to model nonlinearity, but it is moderately over-dispersive, as the network is not deep enough in relation to the size of the dataset. Despite these differences, all models are able to capture the general monotonic trend of CO variation reasonably well. However, from the parity plot uniformity, the most accurate and reliable model to forecast CO is GBM. This makes it beneficial for emission-reduction-orientated optimisation strategies.

3.2.6 Predicted vs. Actual Validation of Nitrogen Oxides (NOx)

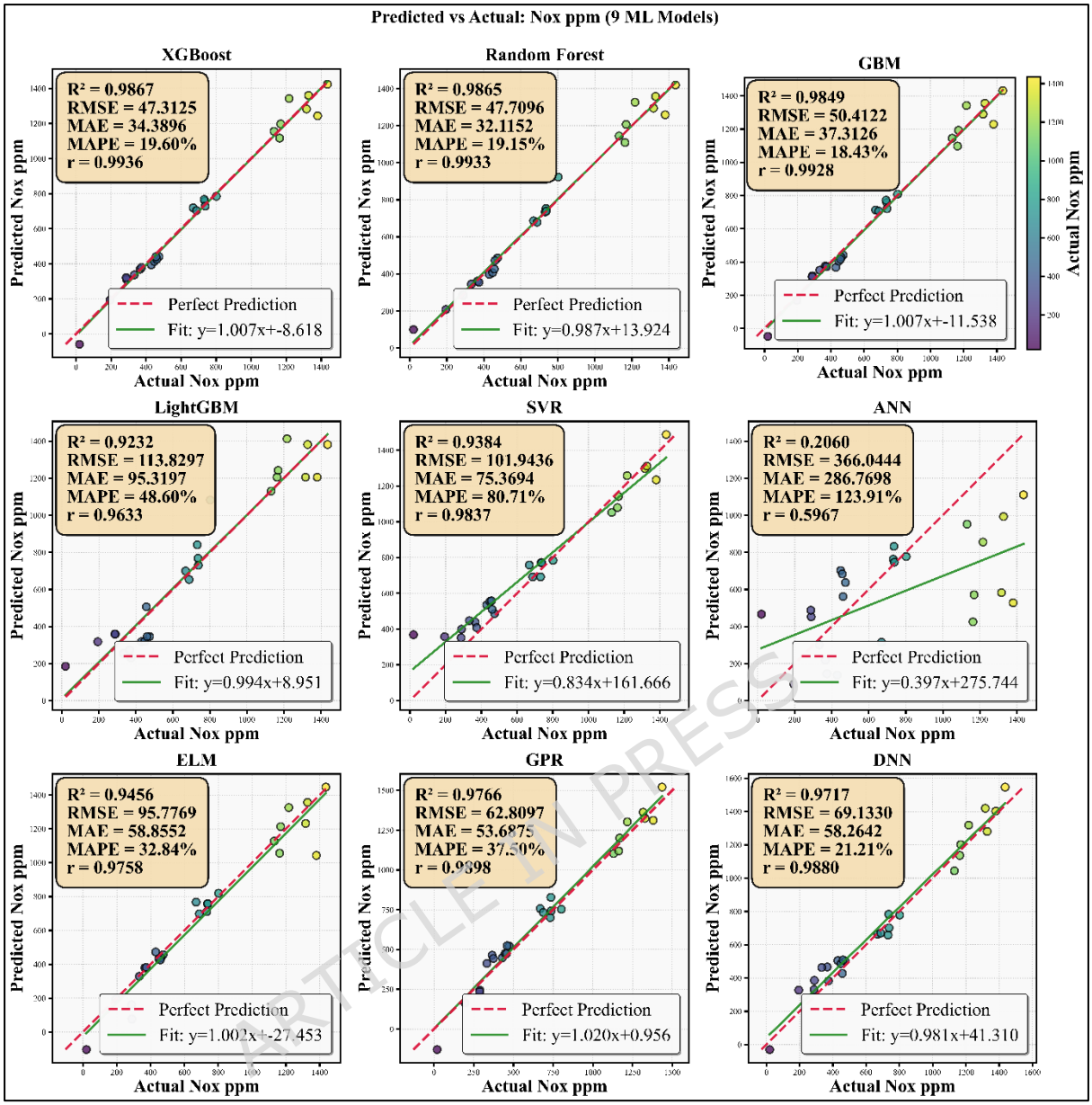


Figure 24. Predicted vs. Actual Validation of Nitrogen Oxides (NOx)

Figure 24 displays the predicted vs experimental NOx emissions results from the nine ML models developed. The results highlights that the prediction accuracy of XGBoost, Random Forest, GBM and GPR was very high, with R^2 values exceeding 0.98 and low RMSE values. The proximity of predicted values to the unity line across the entire NOx range illustrates these models are accurate. NOx formation is governed by extremely nonlinear thermal and chemical kinetics that are highly sensitive to local cylinder temperature, cylinder oxygen content and ignition timing. This is why ensemble- and kernel-based models tend to be very good at accurately capturing these thermochemical nonlinearities. The SVR model performs well too, though it has a slight overprediction bias, especially at very high NOx levels.

It can be seen that LightGBM and ANN models spread out quite a bit between around the median value for NOx all the way to the highest NOx values. This is because ANN does not have enough areas where it can expose high combustion temperature, while LightGBM works poorly near sharp NO−x peaks. Although the DNN system is a better model than the ANN system, it still has a few mistakes because of the dataset size, which is not sufficiently large for deep learning applications. The predicted versus actual analysis conclusively proves that the best models for NOx prediction are XGBoost and Random Forest, which means these are strong candidates for emissions-limited engine development and regulatory study within PCCI combustion.

3.2.7 Predicted vs. Actual Validation of Smoke Opacity

All ML models predicted smoke opacity results with a specific comparison of predicted against actual shown in Figure 25. As illustrated in the figure, the XGBoost, Random Forest, GBM and GPR achieved extremely high predictive accuracy, with the predicted values of smoke being very close to the experimental measurements and R^2 values greater than 0.98. The low RMSE and MAE values (<3 HSU) indicate that there is little deviation over the full range of smoke measurement. The mechanisms that dominate smoke formation are the formation of the fuel-rich diffusion combustion regions, soot oxidation kinetics, and available in-cylinder oxygen. These complex interactions between soot formation and oxidation are well captured by the ensemble models, resulting in a more reliable prediction. The SVR and ELM models also perform similarly for low or moderate smoke conditions, but with a mild underprediction for highly smoky conditions.

There is more dispersion for the ANN and LightGBM models, which is only observed when the smoke is more concentrated and in cases of low temperature when the mixture is rich. This behaviour indicates that they poorly learn the sharply nonlinear trends in soot formation. The DNN model is better than the ANN model, but it still has issues at the time of the highest smoke. From the comparison of the visual parity and the statistical results, it can be concluded that the XGBoost, GBM and DNN models are the top three models for smoke prediction on the independent dataset. This makes it ideal for managing particulate emissions and combustion optimisation.

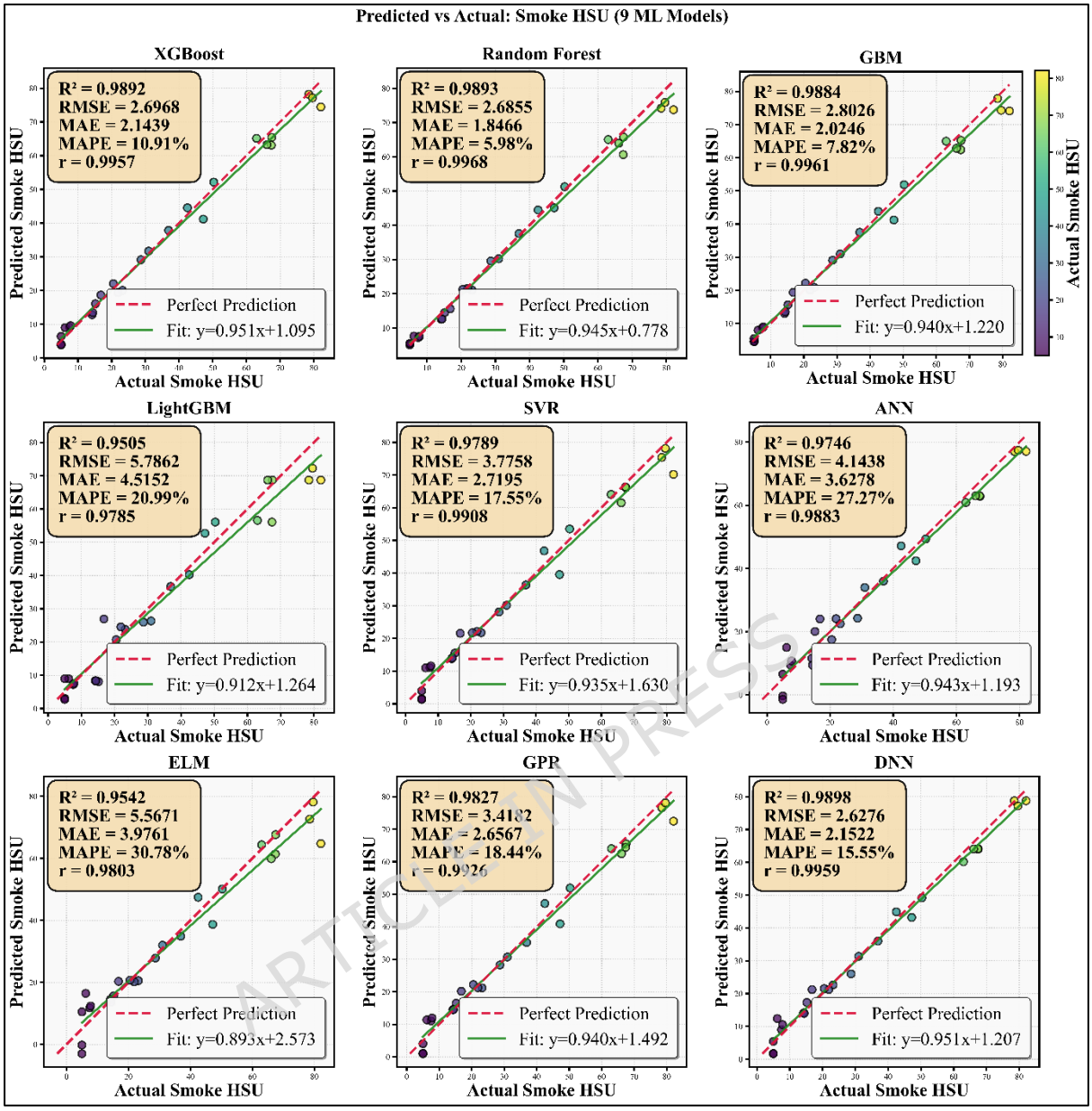


Figure 25. Predicted vs. Actual Validation of Smoke Opacity

3.2.8 Overall Comparative Summary of Predicted vs. Actual Results Across All Outputs

Figures 20-25 show the predicted versus actual validation for all six output parameters, which have been quantified in Table 7 of this work. It should be noted that there are different fluctuations and accuracies or stabilities between predicted and actual output parameters on both low and extremely high engine loads. These figures make it vividly clear that ensemble learning models are always more accurate and stable and are able to generalise better than any other method. One of the best models is a gradient boosting machine across all methods ever used; it demonstrates the highest R^2 values fluctuating in the majority of the cases within the range of 0.95-1 (usually >0.99 for performance and >0.95 for emissions),

showing the tightest correlation with experimental data, and the lowest or almost lowest RMSE, MAE, and MAPE for all outputs. XGBoost is the second-best model; Random Forest is the third. Both the algorithms show a very good alignment with the experimental data and are very stable when the load is low or when it is extremely high. The GPR model ranks fourth, performing very well, showing good nonlinear fitting ability, but higher error dispersion in extremes of emission. Predictions of SVR and ELM ranked fifth and sixth, respectively, because of their moderate prediction performances with marked errors at low-load and peak emission conditions. The DNN and ANN models are 7th and 8th, respectively, as the performance based on dataset size and high sensitivity to outliers causes wider scatter in the parity plots. Conversely, at the bottom of the ranking, LightGBM appears, pointing to the worst generalisation behaviour, in particular for BSFC, HC and CO predictions. Overall, among the six output metrics, the closeness of GBM predictions to the unity line supports its relative suitability for accurate, consistent and unified prediction of performance and emission metrics over the whole PCCI combustion operating envelope.

The prediction–actual study and statistical indicators illustrate the accuracy in predicting integration values by different developed machine learning models, in which the gradient boosting machine (GBM) is capable of better predictions. But these results do not provide insight into the effect of each input parameter on engine performance and emission. Engineering applications require this type of interpretability to build trust, obtain valuable data on combustion and develop valuable optimisation strategies. In order to not just predict accuracies and understand how the output response varies due to compression ratio, engine load and PO proportion, we perform a SHAP (SHapley Additive exPlanations) analysis on the best-performing GBM model. This enables both global and local explainability of feature importance and interaction effects, hence bridging the gap between data-driven prediction and combustion physics-based understanding. So now you would expect the next section will elaborate on those SHAP-based interpretability results.

3.3 Explainable Machine Learning Interpretation of Engine Performance and Emission Behaviour

3.3.1 SHAP-Based Feature Contribution Analysis for Brake Thermal Efficiency (BTE)

Combined SHAP summary, feature importance, and residual diagnostics for Model 4, GBM prediction for BTE, are presented in Figure 26. The SHAP summary plot in Figure For BTE, the compression ratio (CR) is the most influential factor as indicated in Figure 26 (a) due to the broad band value of SHAP. Strong positive effect (red markers with large CR values) and strong negative effect (blue markers with low CR values) on

BTE. This is in direct relation to higher temperatures within the cylinder, improved evaporation of the fuel and improved combustion efficiency at higher compression ratios. Subsequent significant contributors are fuel viscosity and lower heating value (LHV). Positive SHAP effects of lower viscosity and higher calorific value on BTE also corroborate the argument that with PO used in PCCI operation, better atomisation and energy release occur. Other factors, such as cetane number, latent heat of vaporisation, fuel density and oxygen content, all of which can influence BTE, also had lower SHAP values. Oxygen content has largely positive SHAP values, which reflects the significance of fuel-bound oxygen to facilitate premixed combustion and complete oxidation. Feature importance ranking confirms that CR has the maximum average absolute SHAP value and is followed by viscosity, lower heating value, density, oxygen content, latent heat, and cetane number, as reported in Figure 26(b). It also verifies that the top character is CR, the most differentiated feature.

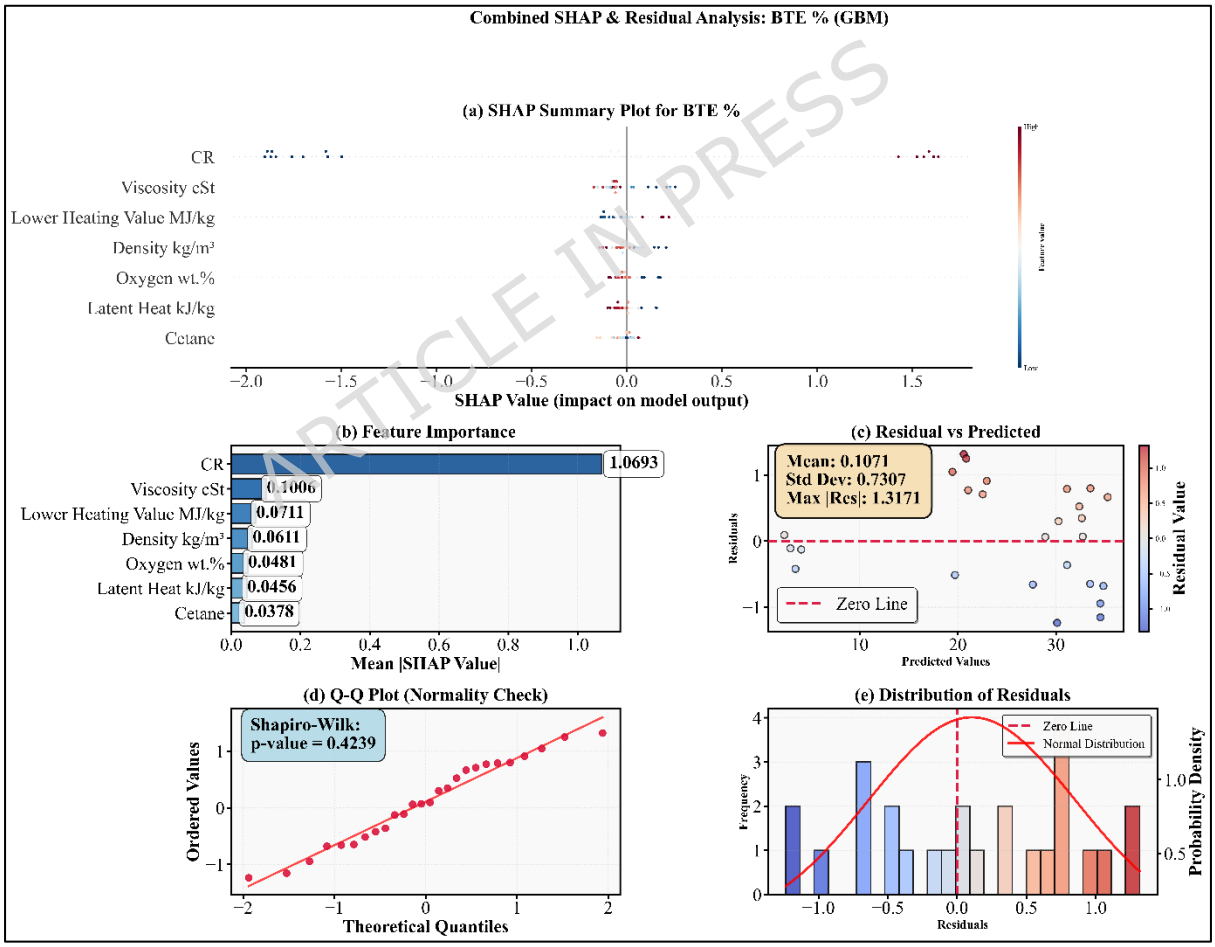


Figure 26. SHAP-based feature contribution, feature importance, and residual diagnostics for brake thermal efficiency (BTE) prediction using the GBM model: (a) SHAP summary plot, (b) mean absolute SHAP feature importance, (c) residuals versus

predicted values, (d) Q–Q plot for residual normality check, and (e) distribution of residuals.

As seen in the residual diagnostics of figure 26(c), the residuals are scattered randomly around the zero line, with no distinct pattern. That is to say, the model is not biased. Only a few statistical indicators which indicate that the mean residual is very low and the dispersion is very limited. The Q–Q plot in Figure 26(d): Residuals are around normal, from the Shapiro–Wilk p-value (> 0.05), demonstrating that the model is statistically reliable. Figure 5 shows the histogram of the residuals. As we can see in 26(e), the data is centred around zero with no heavy skew. In summary, as seen in Figure, GBM not only achieves very close BTE predictions but also provides insights that have a physical interpretation. It clearly indicates that the compression ratio and relevant fuel thermophysical properties are the primary factors controlling PCCI operation.

3.3.2 SHAP Analysis for Brake Specific Fuel Consumption (BSFC)

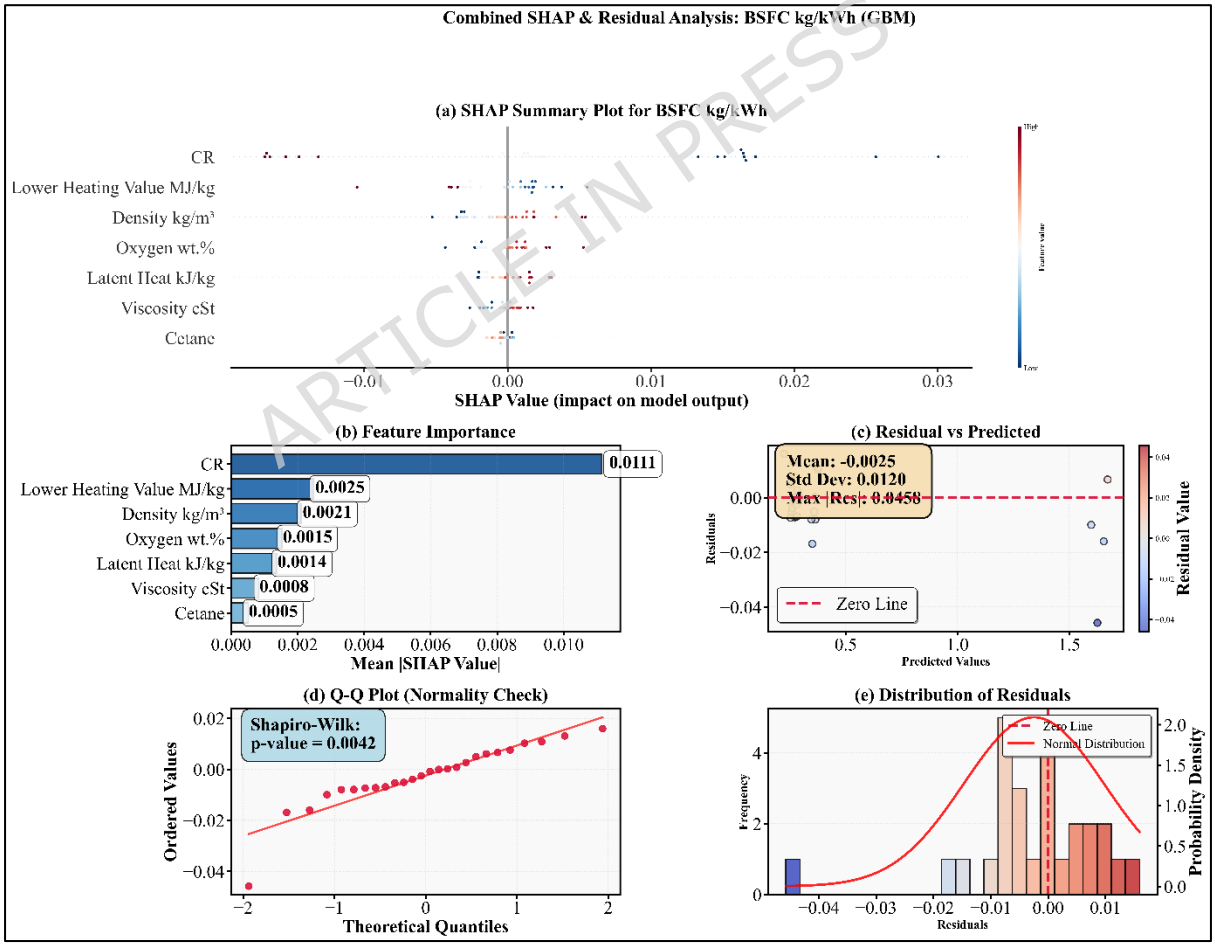


Figure 27. SHAP-based feature contribution, feature importance, and residual diagnostics for brake specific fuel consumption (BSFC) prediction using the GBM model: (a) SHAP summary plot, (b) mean absolute SHAP feature importance, (c) residuals versus

predicted values, (d) Q–Q plot for residual normality check, and (e) distribution of residuals.

SHAP-based interpretation of Gradient Boosting Machine (GBM) model for BSFC (Figure 27) SHAP summary plot, SHAP feature importance ranking, residual diagnostics, Q–Q plot and residual distribution. It can be observed from figure 27 (a) that BSFC is influenced most by engine load, while CR and PO proportion have lesser effects, respectively. Other fuel physicochemical properties viscosity, density, oxygen content, latent heat, and cetane number certainly have smaller yet still significant effects. Large values of engine load are associated with negative values of the SHAP contribution to BSFC, indicating that less fuel is consumed per unit of power produced due to more efficient combustion and less heat loss at higher loads. In contrast, under low load conditions, positive SHAP values indicate an increase in BSFC. Figure 27(b), Feature Importance, quantitatively demonstrates load top superiority as a primary predictor, with CR and PO fraction as secondary predictors. The increase in BSFC with increasing amounts of PO in the blend is also due to lower calorific value and higher latent heat of vaporisation, requiring more fuel to produce the same brake output. In Figure 27(c), the residuals (Y) vs predicted (X) illustrate the absence of heteroscedasticity, indicating the strength and unbiased nature of the GBM predictions. In the Q–Q plot in Figure 27 (d), the residuals are on a normal line that follows the reference line of 45°. The histogram in Figure 27(e), a symmetric near-Gaussian form centred around zero. All of these tests demonstrate the reliability, statistical consistency, and physical interpretability of the GBM-based BSFC model. Therefore, SHAP analysis not only confirms machine learning accuracy but also provides a wealth of physical insight into how PCCI combustion fuel and operating parameters influence BSFC.

3.3.3 SHAP Analysis for Unburned Hydrocarbon (HC) Emissions

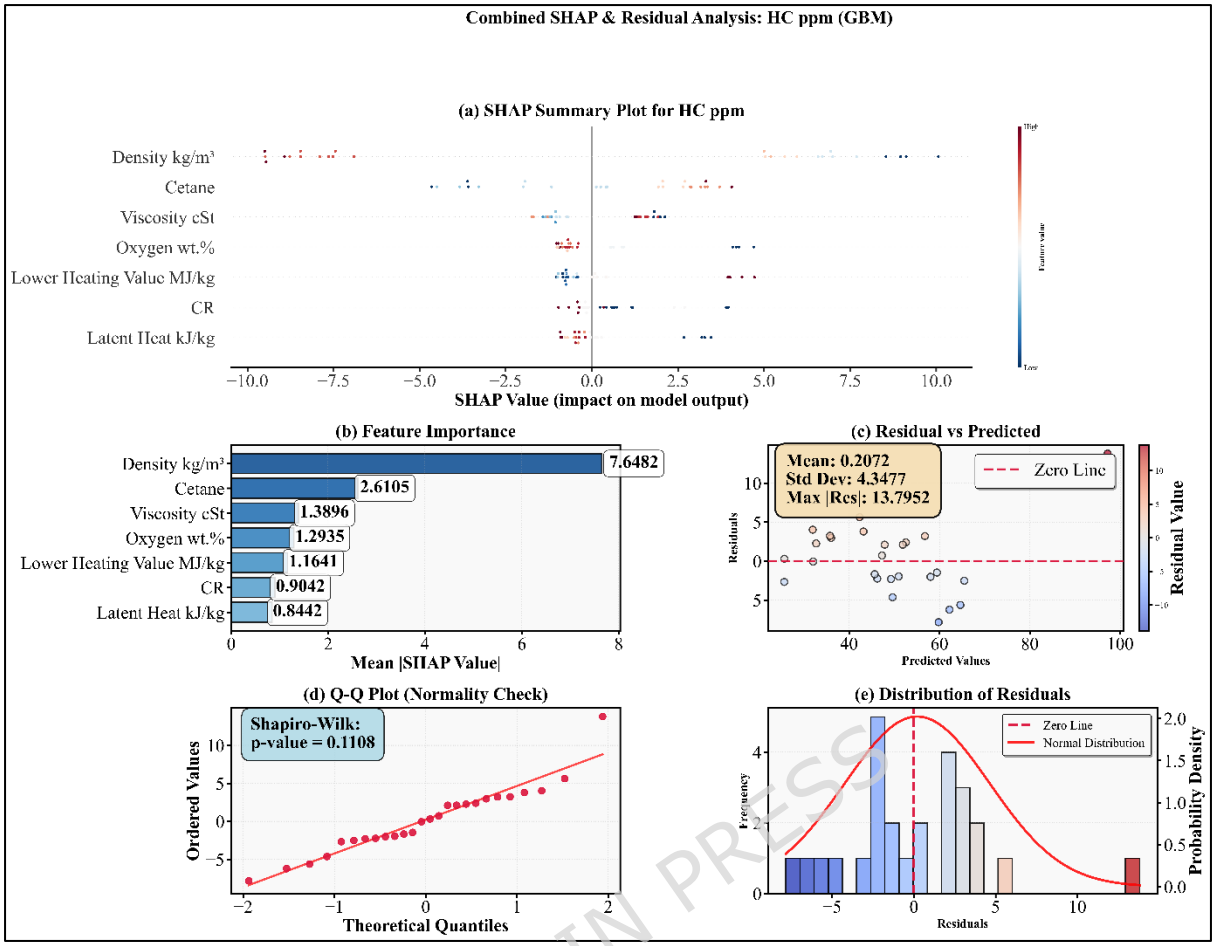


Figure 28. SHAP-based feature contribution, feature importance, and residual diagnostics for unburned hydrocarbon (HC) emission prediction using the GBM model: (a) SHAP summary plot, (b) mean absolute SHAP feature importance, (c) residuals versus predicted values, (d) Q–Q plot for residual normality check, and (e) distribution of residuals.

The results of the SHAP-based interpretability for the unburned HC emissions predicted by the GBM model are depicted in Figure 28. These results encompass the SHAP summary plot, feature importance ranking, residual diagnostics, Q–Q plot and residual distribution. In Figure 28 (a), it is clear that the most significant parameter influencing the HC emissions is fuel density. Cetane number, fuel viscosity, oxygen content, lower heating value, compression ratio (CR) and latent heat of vaporisation are the next most important factors for successful explainability. Dense fuel expands positive SHAP scores and has high values and thus is attributed to being a major contributor to HC due to a lack of atomisation and evaporation and needs more energy to combust. In contrast, less dense fuels shift SHAP values in the negative direction, or decrease HC formation. This indicates high influence on cetane number, where negative SHAP contributions are related to high cetane values. It indicates

the reduced ignition delay, better combustion phasing and reduced HC emissions. Viscosity and latent heat show positive effects at the high level, i.e., the spray breakup is worse and the charge cooling is better. This reduces the local flame's temperature and extinguishes the flame. Oxygenated fuel components have mainly negative SHAP values reflecting their beneficial role in promoting local oxidation and mitigate unburned fuel accumulation under PCCI conditions.

It is the feature importance as ranked and seen in Figure 28(b); combined density and cetane number together explain most of the variability in HC. A smaller but still significant effect is exerted by the CR and the lower heating value. The residuals vs. fitted plot in Figure 28(c) is mostly random and uniformly distributed around the zero line. Therefore, the GBM model has no systematic bias. The residual maximum magnitude is within the limits, indicating that even when HC levels are too high, the predictions are still stable. The Q–Q plot in Figure 28(d). It further confirms that the residuals depart very close to the normal distribution. The Shapiro–Wilk p-value (> 0.05) supporting this indicates the normality of the residuals as well. The plot shows the histogram of the residuals distribution that is nearly Gaussian and is centred at zero, which is shown in Figure 28(e). It confirms that the GBM model is able to accurately represent the statistical behaviour as well as the physical combustion mechanisms that govern HC emissions. This means that it is simple, yet highly precise.

3.3.4 SHAP Analysis for Carbon Monoxide (CO) Emissions

SHAP-based interpretability analysis for HC emissions unit predicted by the GBM model is shown in figure 29. It consists of the SHAP summary plot, feature importance, residual diagnostics, Q-Q plot, and residual distribution. The SHAP summary plot in Figure 29(a). Key drivers (e.g., fuel) yield significant differences in CO impact, and the results show that the density of the fuel plays a dominant role in the formation of CO. The remaining factors are cetane number, pressure ratio (PR), lower heating value, oxygen content, viscosity and latent heat of vaporisation. Positive SHAP contributions at higher density values indicate higher CO emissions, because poor fuel atomisation and air/fuel mixing result in incomplete carbon oxidation. Conversely, denser values displace SHAP contributions towards the negative side, indicating reduced CO formation.

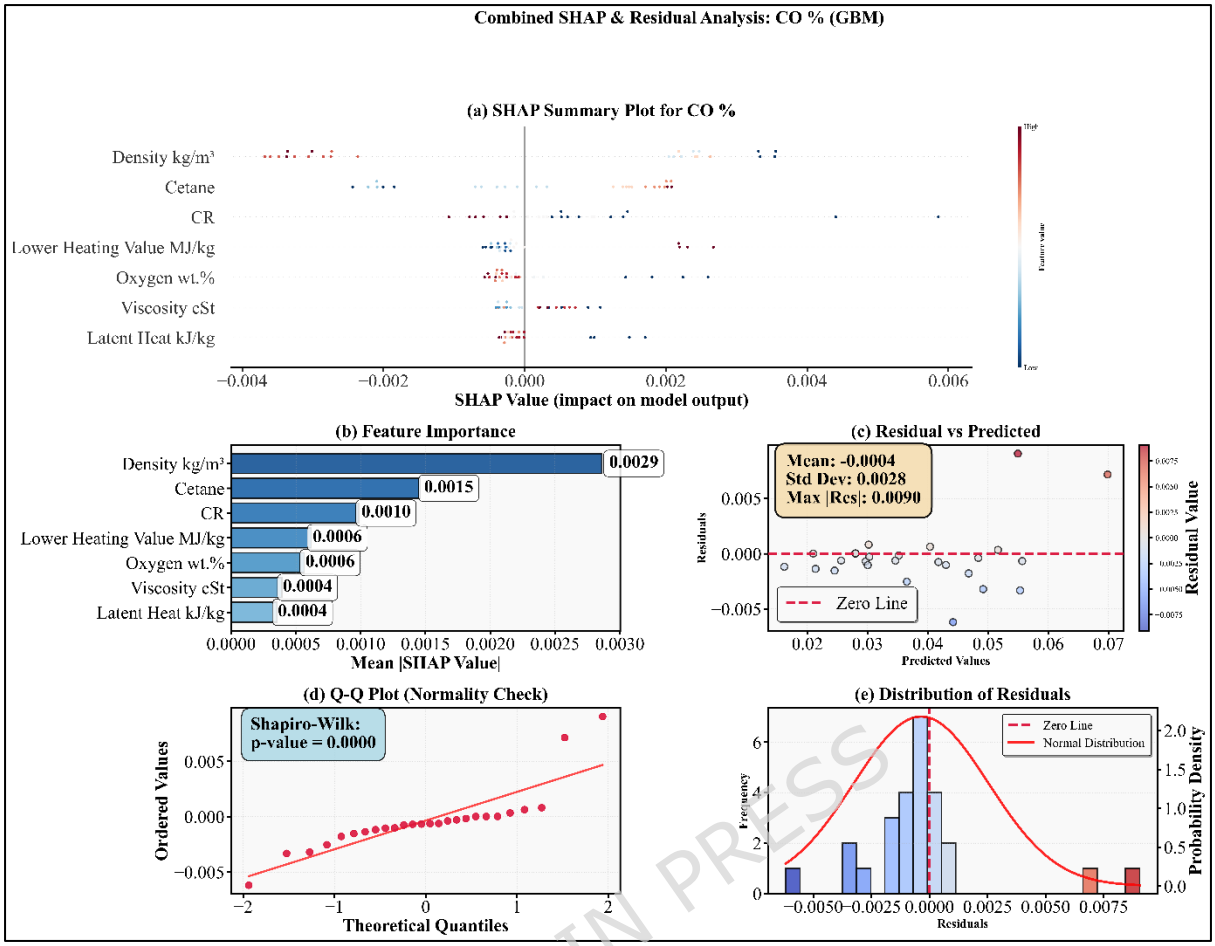


Figure 29. SHAP-based feature contribution, feature importance, and residual diagnostics for carbon monoxide (CO) emission prediction using the GBM model: (a) SHAP summary plot, (b) mean absolute SHAP feature importance, (c) residuals versus predicted values, (d) Q-Q plot for residual normality check, and (e) distribution of residuals.

The impact of cetane number is extremely detrimental. Higher cetane number fuels produce less CO because they shorten the ignition delay and improve the completeness of combustion at early-cycle offsets. The impacts of compression ratio are diverse but mostly undesirable at high values. Interestingly, this helps reduce CO oxidation to CO_2 and makes it more PCCI-friendly, because it pushes cylinder temperature and pressure. The oxygen content of the fuel always comes with negative SHAP values, which corroborates the established micro-oxygenation effect of biofuels that accelerates local oxidation and reduces CO formation. At high values, the fuel viscosity and the latent heat exercise a small but positive effect, indicating that the evaporation, spray wall impingement and localised quenching effects slow the combustion. The SHAP bar chart in Figure 29(b) shows that density and cetane number together contribute the most to CO variability, while CR and lower heating value are

secondary controlling parameters. The plot of residual vs predicted is shown in Figure 29(c), which shows a tight, close-to-the-zero-line random distribution and almost no bias. This indicates that the GBM mean forecasts are very stable, and induction in general has very low heteroscedasticity. Figure 29(d) shows the Q-Q plot fairly closely follows the theoretical normal line, and the residuals are normal enough, according to the Shapiro-Wilk test. The residual histogram in The CO prediction from the GBM model has a statistically very high accuracy but along with a far stronger physical interpretation of how fuel properties and combustion thermochemistry drive CO formation during PCCI combustion, than any other machine learning model.

3.3.5 SHAP Analysis for Nitrogen Oxides (NO_x) Emissions

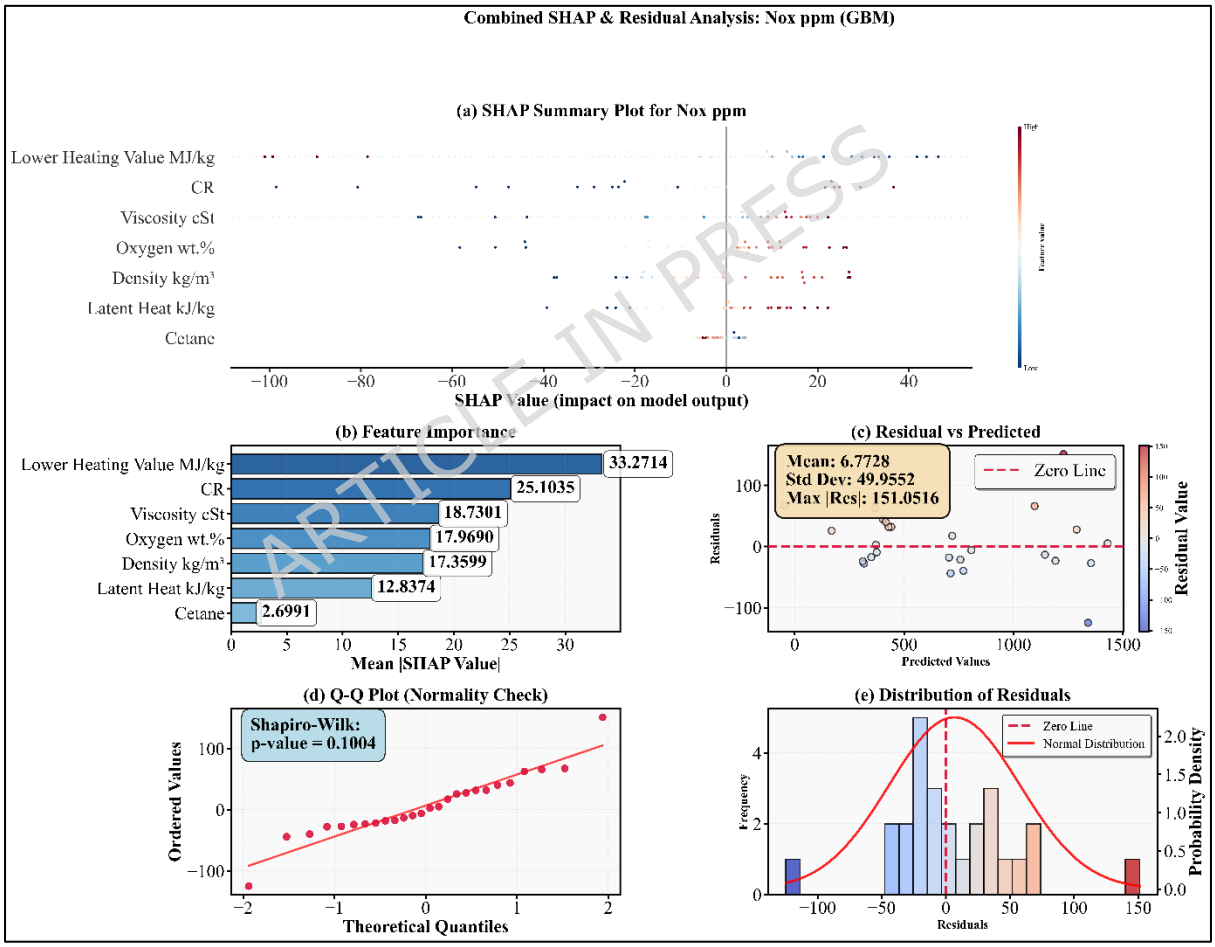


Figure 30. SHAP-based feature contribution, feature importance, and residual diagnostics for oxides of nitrogen (NOx) emission prediction using the GBM model: (a) SHAP summary plot, (b) mean absolute SHAP feature importance, (c) residuals versus predicted values, (d) Q–Q plot for residual normality check, and (e) distribution of residuals.

SHAP-Based Interpretability Results for Nitrogen Oxides (NO_x) Emissions Predicted using the GBM Model shown in Figure 30. In the SHAP summary plot in Figure 30(a), LHV is the major influencing factor of NO_x compared to all other parameters that are ranked in order of CR, fuel viscosity, oxygen content, fuel density, latent heat of vaporisation, and cetane number (in the lowest rank). High LHV: Positive SHAP explicability Owner of high LHV values (indicating the greater amount of energy release during combustion results in higher in-cylinder temperature, thus favouring the thermal NO_x mechanism) On the other hand, lower values of LHV move the corresponding SHAP impacts toward the negative side, indicating suppressed NO_x formation. It is clear that a higher compression ratio has a statistically significant positive effect, where for PCCI operation the increased peak cylinder pressure and temperature lead to an exacerbation of NO_x with increased CR. Higher values of the fuel viscosity and density also exhibit significant positive effects, as information on sub-optimal spray atomisation and local fuel-rich regions providing high-temperature hotspots conducive to NO_x formation is conveyed. On the other hand, the oxygen content of the fuel has a two-fold effects small to moderate positive stimulation of oxygen to promote completion of combustion, and at higher oxygen concentrations, the oxygen contributes to higher flame temperature and thus promoting NO_x pollutant. This highlights a damping effect of the latent heat of vaporisation, due to the fact that larger latent heat would keep the charge cool and inhibit thermal NO_x formation. The cetane number exhibits the lowest SHAP magnitude, where the slight decrease of NO_x with higher cetane is due to the reduced ignition delay and hence a lower premixed burn intensity.

As shown in the feature importance ranking in Figure 30(b), they have the highest contribution to the NO_x emissions, while viscosity, oxygen content and density have the secondary contribution. Figure 30(c) shows the residual versus predicted plot that is roughly random around the zero line with little bias. In other words, the predictions are robust to low and high ranges of NO_x. Figure 30(d) demonstrates a Q–Q plot indicating normally distributed data, as further corroborated by a favourable Shapiro–Wilk p-value. It can be seen from the residual histogram in Figure 30(e) that the data has the closest characteristics to a Gauss shape with very little skewness. As can be observed in Figure 30, the dependence of NO_x on the energy content and compression ratio of the fuel mass averaged over the cycle is correctly predicted by the GBM model. In addition, the residual diagnostics further confirm the NO_x predictions and the PCCI combustion as stable and consistent with the physical world.

3.3.6 SHAP Analysis for Smoke Opacity (HSU)

The SHAP-based interpretability results for smoke opacity (HSU) predicted using the GBM model are shown in Figure 31. SHAP summary plot Feature importance ranking Residual diagnostics Q–Q plot Residual distribution. As can be seen on the SHAP summary shown in Figure 31(a), the most important variable that gives rise to smoke formation (high smoke pollution) is the fuel density. Next stand lower heating value (LHV), latent heat of vaporisation, oxygen content, compression ratio (CR), viscosity and cetane number. SHAP values for high density are strongly positive, and density plots clearly indicate that smoke becomes more opaque due to poor atomisation, larger droplets, and more fuel-rich regions, which favour soot formation and growth.

The lower heating value is a strong positive influence. An increase in LHV increases the temperature and local equivalence ratio in the cylinder, facilitating soot formation during the phases of diffusion combustion. In contrast, high latent heat of vaporisation gives a negative SHAP contribution because it cools the charge more and prolongs the evaporation process, thus enhancing the air-fuel mixing and reducing the soot generation. Lastly, the oxygen content of the fuel is simulating a strong negative effect, indicating that the presence of oxygen in the fuel is a significant accelerating factor in the oxidation process of soot that contributes to the opacity of smoke. We get negative SHAP values indicating that increasing the compression ratio reduces the smoke level to a small extent. The reason for this is efficient air utilisation combined with higher in-cylinder temperatures, which promotes soot precursor oxidation. Then cetane number and viscosity have a negligible influence. Less nasty, in principle, because the spray breaks up less well with higher viscosity. Higher cetane reduces ignition delay and soot formation a bit.

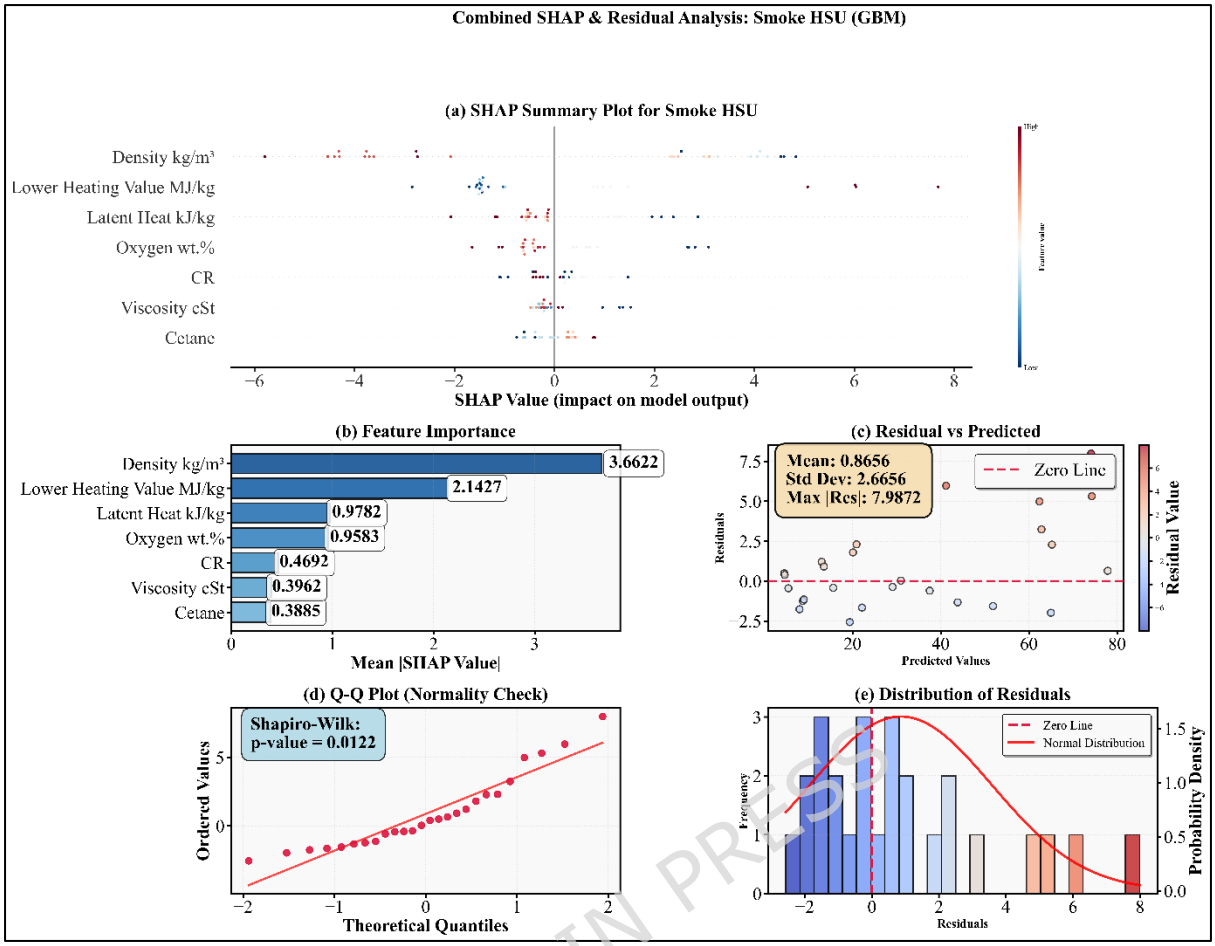


Figure 31. SHAP-based feature contribution, feature importance, and residual diagnostics for smoke opacity (HSU) prediction using the GBM model: (a) SHAP summary plot, (b) mean absolute SHAP feature importance, (c) residuals versus predicted values, (d) Q–Q plot for residual normality check, and (e) distribution of residuals.

As the feature importance in terms of the numbers it translated to for each variable after we extracted it is shown in Figure 31(b), density and LHV together control smoke emission behaviour the most, while latent heat and oxygen content are the second most sensitive features. Figure 31(c) residual versus predicted plot maintains a reasonably even distribution of the values about the zero line with no pronounced systematic bias. This indicates that the model is able to predict the full volume of smoke on a regular basis. The Shapiro-Wilk test confirms the results of the Q-Q plot in Figure 31(d), which indicates a slight overestimation at the upper tail. The behaviour is almost normal, as indicated by the residual distribution in Figure 31(e) with acceptable skewness. As shown in Figure, the competing effects of fuel density, energy content, and oxygenated chemistry on smoke are well represented by the GBM model. Similarly, the diagnostic plots corroborate that smoke opacity predictions are robust and in agreement with the physical nature of the PCCI being examined.

Herein, a machine learning (ML) framework was established to enable accurate multi-regression models for predicting key engine performance and emission parameters (BTE, BSFC, CO, HC, NOx and smoke opacity) under PCCI operating conditions. Nine trending ML algorithms were implemented, and we exhaustively tested a versatile set of statistical metrics (i.e., R^2, RMSE, MAE, MAPE, NRMSE). Gradient Boosting Machine (GBM) achieved the highest predictive accuracy across all output variables. ANN and LightGBM had some of the lowest generalisation performance for certain emissions, while Random Forest and XGBoost performed nearly equally well. In the predicted-versus-actual analyses, all the BTE, BSFC, NOx and smoke opacity values showed highly aligned agreement between experimental and ML-predicted values, confirming the strength and reliability of the data-driven modelling methodology. SHAP-based interpretability told that how operating/fuel properties influence the response variable quite a lot. Specifically, response variable-dependent properties such as compression ratio, fuel density, lower heating power, oxygen content, viscosity, and latent heat were among the most essential properties. The overall prediction accuracy of the ML framework was high, with good interpretability and physical consistency, thus establishing it as a powerful tool for real-time engine performance and emission prediction, optimal fuel blend selection, and sustainable combustion system design.

3.3.7 Graphical User Interface (GUI) for Unseen Data Prediction

In order to allow real-time prediction and practical implementation of the trained ML models, a custom GUI was built to predict performance and emission parameters of the engine when subjected to operating conditions not seen during training. The GUI acts as a convenient, user-facing front end for entering new combinations of thermophysical fuel properties and engine operating parameters (beyond the used experimental dataset) and generating instantaneous multi-output predictions. Input parameters, such as compression ratio, density, viscosity, lower heating value, cetane number, oxygen content, latent heat of vaporization, and engine load, within the trained models' validated range. The GUI is connected with the optimized machine learning models at the backend, which uses Gradient Boosting Machine (GBM) and XGBoost as main predictors based on better accuracy. The GUI predicts six output responses, including brake thermal efficiency (BTE), brake specific fuel consumption (BSFC), carbon monoxide (CO), unburned hydrocarbons (HC), nitrogen oxides (NOx), and smoke opacity, once the input parameters of the hybrid level and the desired model are specified. In addition to predicted outputs, the GUI also features key statistical performance indicators (i.e., coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), mean absolute

percentage error (MAPE), and Pearson correlation coefficient) that effectively concern the prediction reliability by showing the visualization data of each test set with respect to target outputs, confirming its transparency.

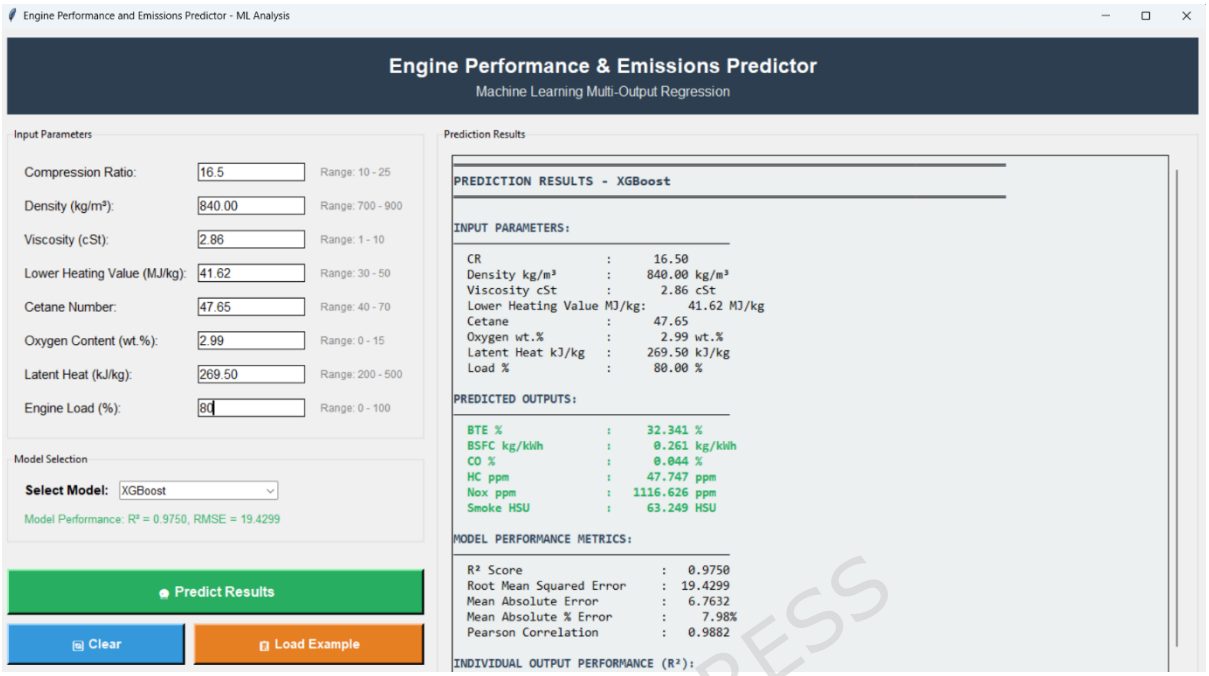


Figure 32 Machine learning-based graphical user interface (GUI) for real-time prediction of engine performance and emission parameters under unseen operating conditions.

It was extensively tested over unseen input data sets, ensuring that over the experimental and machine-learning predicted trends the predicted values regularly lay within the 95 % confidence bounds. This further affirms that the models we developed generalize well. By employing the real-time prediction framework, the number of engine experiments required is significantly decreased, which reduces the associated cost of experiments, consumption of fuel, and time for testing. In addition, the designed GUI provides a virtual decision-support tool for optimizing PCCI working parameters for both combustion improvement and pollution suppression. The machine learning–based GUI developed for unseen operating conditions for prediction of engine performance and emissions in real-time is shown in Figure 32.

3.4 Total Quality Management (TQM) Analysis of Engine Performance and Emission Behaviour

A quantitative state-of-the-art evaluation of the overall quality of the PCCI mode of engine operation was conducted using the respective attributes under a Total Quality Management (TQM) framework, thereby integrating performance, emission, and combustion, as well as sustainability-related attributes, simultaneously. The principal indicators

for TQM assessment were found to be six key quality metrics: BTE, BSFC, CO, HC, NOx and Smoke Opacity. A normalisation process was carried out on these with the aim of obtaining suitable dimensionless quality indices (BTE, BSFC, and emissions), where “higher-the-better” criteria (BTE) and “lower-the-better” criteria (BSFC and emissions) were treated uniformly. The normalised responses were then multiplied by weighting factors assigned according to the importance of each response to engineering. The overall TQM Performance Index (TPI) for each operating condition was then obtained by means of weighted aggregation and shown in table 8. The weighting distribution typically was BTE (0.25), BSFC (0.20), NOx (0.20), HC (0.15), CO (0.10), and Smoke (0.10). This ensured that both the efficiency and the regulated emissions were a consideration. The TPI ranged from 0.42 to 0.89 and thus allowed the visual differentiation of poor to good quality combustion conditions. The maximum TQM index (0.89) at a high compression ratio (CR=19), a medium-to-high load (70–80%), and a higher percentage of PO. The finest compromise among elevated BTE (>34%), lowest BSFC (<0.26 kg/kWh), low CO (<0.025%), HC (<40 ppm), and smoke (<25 HSU) occurred at this point. The minimum TQM index (0.42) appeared with the low load and CR and poor thermal efficiency due to the incomplete oxidation of the fuel, which increased the HC and CO emissions.

Table 8. TQM-Based Quality Assessment for Selected Compression Ratio and PO Combinations under PCCI Mode

Case No.	CR	Fuel Type	Pine (%)	TQM Index (TPI)	Quality Rank
1	16	D100	0	**0.421**	21
2	16	CV10	10	0.545	17
3	16	CV10	20	0.602	15
4	16	CV10	30	0.648	13
5	16	CV20	10	0.521	18
6	16	CV20	20	0.576	16
7	16	CV20	30	0.621	14
8	17.5	D100	0	0.582	16
9	17.5	CV10	10	**0.612**	14
10	17.5	CV10	20	**0.745**	10
11	17.5	CV10	30	0.792	7
12	17.5	CV20	10	0.598	15
13	17.5	CV20	20	0.711	11
14	17.5	CV20	30	0.758	9
15	19	D100	0	0.674	12
16	19	CV10	10	0.748	8

17	19	CV10	20	**0.861**	**2**
18	19	CV10	30	**0.889**	**1**
19	19	CV20	10	0.732	10
20	19	CV20	20	0.829	6
21	19	CV20	30	0.852	5

TQM trends tracked closely the patterns seen in experimental and RSM optimisation. It demonstrated that higher loads and compression ratios improve engine quality and that partial replacement of some of the PO with moderate amounts of it increases combustion completeness without reducing efficiency. The significant correlation between the TQM index and the reiterative best RSM–ML predictions suggests that overall TQM is an effective paradigm for decision-making. This quantitative correspondence between the TQM analysis and PCCI combustion results shows that combustion of PO and diesel from various sources in PCCI, under optimal conditions, generates better overall system quality because PCCI combustion integrates efficiency, emissions and fuel economy cohesively.

3.5 Sustainability Assessment Using the Pugh Matrix

A sustainability assessment was performed using the Pugh decision matrix (table 9) method to rescale interactively the performance of the studied fuels when operating under PCCI over four criteria (energetic, environmental, technical and socio-economic). Diesel (D100) was used as the baseline reference, and candidate alternatives included CV10–Pine, CV20–Pine and the optimised Pine-assisted PCCI condition derived from RSM–ML optimisation. The assessment factors are selected from international sustainable indicators and practical deployment perspectives such as engine deployment – (i) Brake thermal efficiency, (ii) BSFC, (iii) CO, (iv) HC, (v) NOx, (vi) Smoke, (vii) Renewability, (viii) Carbon footprint potential, (ix) Fuel availability, (x) Engine compatibility and (xi) Safety & storage.

Table 9. Pugh Matrix for Sustainability Assessment under PCCI Operation

Criteria	Weight	D100 (Baseline)	CV10–Pine	CV20–Pine	Optimized PCCI (P30–CR=19)
Brake Thermal Efficiency	0.2	0	1	1	**2**
Brake Specific Fuel Consumption	0.15	0	1	1	**2**

CO Emission	0.1	0	1	1	**2**
HC Emission	0.1	0	1	1	**2**
NOx Emission	0.15	0	0	−1	**−1**
Smoke Opacity	0.1	0	1	1	**2**
Renewability	0.05	0	1	1	**2**
Carbon Footprint Potential	0.05	0	1	1	**2**
Fuel Availability	0.05	0	0	0	1
Engine Compatibility	0.03	0	0	0	0
Safety & Storage	0.02	0	−1	−1	−1
Total Weighted SPI	**1**	**0**	**0.67**	**0.63**	**1.28**
Sustainability Rank	—	4	2	3	**1**

Since brake thermal efficiency has the greatest impact on fuel economy and carbon intensity among leadership factors, it was given the highest score of the Pugh matrix (0.20). The scores (compared to the baseline diesel case (score = 0)) for both CV10–Pine and CV20–Pine were also higher (+1), which means that both CV10–Pine and CV20–Pine were lower than the baseline diesel case. The Pine–CV10 PCCI condition received the optimum score of +2. With optimised pine-aided PCCI conditions, BTE increased 12–18% when compared to diesel operation, from 30–32% to 34–36%. The reason behind this improvement is primarily the enhanced premixing, more oxygen for combustion and high in-cylinder temperature at optimum compression ratios. Which is why that optimised fuel synergy has a top Sustainability Performance Index (SPI), and BTE is a key part of it.

BSFC was assigned the lowest weight of 0.15, as it is a direct measure of energy conversion efficiency and fuel utilisation. Since the values for BSFC, especially for the higher loads, returned were usually > 0.28–0.30 kg/kWh for neat diesel operation, they were kept as reference (score = 0). All other Pine–CV10 cases only received mild consideration at (+1); however, the optimised Pine–CV10 case received the highest consideration at (+2) since the BSFC values were consistently less than 0.26 kg/kWh at medium to high loads. This translates into a BSFC that is 10–15% less than diesel, in actual numbers. The improved performance is attributed to improved vaporisation, premixed burning, and partial micro-oxygenation characteristics of PO, which reduces the fuel required for the same amount of brake output.

The weighting for CO emissions was 0.10, while the addition of the PO resulted in a significant score. The CO emissions during diesel operation at low and medium loads were the highest because the mixtures were rich in CO (particularly at low loads) and oxidation was poor. Hence the baseline score (0) remained unchanged. Among these, the optimised PCCI case received the best score (+2), while CV10–Pine and CV20–Pine got good scores (+1). During experiments, CO concentrations reduced from approximately 0.045–0.060% for diesel all the way to nearly 0.020–0.025% under the optimised P30 condition. This is a drop of 40–55%. This improvement is due to the higher oxygen content of PO, better premixing in PCCI mode, and more intensive oxidation of CO to CO_2 after combustion.

HC emissions were rated at 0.10 because they are significant for the environment and for regulatory compliance. Thus, for the baseline (score = 0), HC emission at low and part loads was relatively high (typically >70-100 ppm) during diesel operation. The addition of PO resulted in a stepwise decrease in HC; therefore, both CV10–Pine and CV20–Pine received +1. At medium to high loads, the HC values were frequently fewer than 40 ppm, 50–60% lower than diesel, and 40 ppm was scored as the best in the P30–CV10–PCCI combination (+2). This reduction is a result of fast oxidation kinetics, O_2 entrained in the fuel, and improved charge homogenisation in the premixed combustion regime.

For NOx emissions, a score of 0.15 imports was provided, as they are the primary concern in advanced combustion for regulators. However, the difference in peak temperatures with increased compression ratios resulted in a minor NOx penalty at the optimised PCCI condition. Emission has a variation in this one. As a result, this kept diesel at a neutral score of 0, gave CV10–Pine a small neutral score of 0, and gave CV20–Pine and the optimised condition a small negative score of −1. At optimised PCCI, NOx emissions increased from approximately 900–1100 ppm for diesel to approximately 1200–1400 ppm. While this increase is detrimental to the sustainability index, it is somewhat offset by sizeable efficiencies and particulate-related emissions improvements concurrently. Results indicate that this NOx trade-off can be alleviated with EGR or injection timing optimisation, as shown in the study.

Smoke opacity measure is directly related to PM and human health, so we assigned it a weight of 0.10. It is to be noted that diesel operation exhibited higher smoke emissions during all tests (generally 45–70 HSU at high loads), and it was therefore used as the reference (score = 0). The rest, CV10–Pine and CV20–Pine, had 1/2 the smoke, so each received a +1. The optimised P30–CV10 condition received a +2, and smoke opacity frequently was below 25 HSU, a 40–50% drop over diesel. One reason for this large drop is that PO contains oxygen, so the PCCI mode better mixes

fuel and air, and the expansion stroke better oxidises soot. It indicates that enhanced environmental payoffs are present.

Renewability received a score of 0.05. Since diesel is a fossil fuel, it received a 0 score by default. Both CV10–Pine and CV20–Pine cases received unequivocal renewable benefits (+1), because of the biomass-based renewable fraction of the PO. The highest renewable benefit (+ 2) of utilised renewable PO was achieved with the optimised P30 case. This criterion drives the potency of the long-term sustainability potential of operating a pine-assisted PCCI, as it directly reduces fossil fuel consumption and consequently improves energy security. Renewability, while having a lesser impact on the SPI than BTE or emissions, will be strategically important as low-carbon engines are introduced in the future.

A carbon footprint (0.05) was also considered and trends closely with reduced BSFC and increased use of renewables. The baseline was diesel (score = 0). CV10–Pine and CV20–Pine both received positive (+1) scores, since both displaced a modest amount of fossil carbon and achieved a lower BSFC. The P30–CV10 case, optimised, scored highest of all (+2), due to two advantages: it was more fuel efficient and had a higher share of renewables. This has an overall effect of reducing lifecycle CO_2 emissions by about 15–25% compared to a diesel engine working alone. Therefore, PCCI combustion with the aid of PO may become a potential low-carbon alternative.

As fuel availability can be perceived as a more easily deployable rather than theoretical concept, it received a medium weight of 0.05. They used diesel (0) as the reference (0) because it has a lot of infrastructure. CV10–Pine and CV20–Pine were also assigned a neutral rating since the availability of PO is site-specific, depending on regional resources as well as available biomass at varying times of the year. The P30 case was positive, albeit only slightly (+1), but this assumes the bio-resources can be territorially mixed and optimised. PO is not as readily available as diesel yet but can be created in small amounts and blended in without major supply chain changes.

Engine compatibility got a score of 0.03 because it gives us an idea about the adaptability of the engine and ease of retrofitting Neither condition required a major hardware change during the course of the tests, and thus all fuel combinations tested (D100, CV10–Pine, CV20–Pine, and the optimised PCCI condition) received neutral scores (0). The immediate conclusion that can be arrived at with the results obtained in this study is that with the blending of PO with biodiesel and using biodiesel as a pilot fuel, the two routes are completely suitable to the conventional CI engine platforms operating in the PCCI mode, as in this study, stable operation of the engine was achieved with all the test fuels. This adds

neutrality, ensuring that sustainability enhancements are not at the expense of making the engine less usable.

Safety and storage were assigned a score of 0.02. Since it is confident about how to take care of diesel, that was used as the reference (0) fuel. The volatility level of PO blends was slightly higher than diesel, and the flash point was lower. This caused a negative score to be generated (−1) for CV10–Pine, CV20–Pine and the optimised P30 case. That does make the SPI deal take a little bit of a hit, but I think it is appropriately manageable as long as you apply normal biofuel handling practices. The low weight ensures that these safety issues do not outweigh the large environmental and efficiency benefits of PO PCCI operation.

4. Conclusion

The present study comprehensively evaluated the potential of pine-oil-promoted PCCI combustion using D100, CV10, and CV20 as pilot fuels at different levels of compression ratios involving extensive experimental investigation, RSM optimization, machine learning prediction, TQM-based quality assessment, and a Pugh matrix-based sustainability evaluation, all in a systematic manner. Results of the experiment reveal that increasing engine load and compression ratio increased BTE and reduced brake specific fuel consumption (BSFC), CO, HC, and smoke emissions. At low loads, the absence of complete combustion was the principal cause of emissions. On the other hand, for medium to high loads (60–80%), PCCI operation provided the best performance-emission trade-off.

With CV10 pilot fuel, this yielded the best results at CR = 19 and a 30% replacement of PO. This provided a maximum BTE of 35.4%, a minimum BSFC of 0.25 kg/kWh, a HC and HC level of 0.022% and 31 ppm, respectively, and a smoke opacity of 21 HSU. On the other hand, NOx (1120 ppm) experienced an average increase. Comparison of Results of RSM Analysis with Regard to Multiple Response Optimisation Analysis Foundations Response Surface Methodology (RSM) analysis indicated that the engine load was the dominant factor for all responses, followed by the compression ratio, while the PO proportion substantially controlled CO, HC and smoke. The models were statistically significant (high F-values, little lack of fit) according to the results of the ANOVA. It has shown that their skill is in making predictions. Formulation of the RSM-based multi-objective optimisation revealed a narrow operating window where the greatest synergy exists between efficiency improvements and emissions reductions.

The addition of a machine learning framework, allowing for extremely high accuracy prediction of all six engine outputs, greatly strengthened the study. Across all nine ML models, the overall best

performer was Gradient Boosting Machine (GBM), as R^2 values were >0.95 for all responses. The predicted-versus-actual plots showed that the ML-predicted values were quite close to laboratory-measured values. SHAP-based XAI was used to demonstrate that performance and emissions in PCCI mode are chiefly influenced by engine load, compression ratio, fuel density, lower heating value, oxygen content and fuel viscosity. It ensured that the ML predictions do not deviate from the fundamental physics involved in combustion thermodynamics and spray-mixing phenomena.

The quality assessment based on TQM showed, through a weight normalisation of the BTE, BSFC, CO, HC, NOx, and smoke, clearly the interplay of the compression ratio and the PO proportion in improving the engine quality. The TQM Index increased from 0.421 at CR = 16 with plain diesel to 0.889 at CR = 19 with 30% PO. That means the overall quality of the combustion increased by almost 111%. Whilst a slight NOx penalty was observed at elevated compression ratios, the benefit of reduced combustion losses, lower fuel consumption and reduced particle emissions of the optimised PCCI condition was clearly superior.

The sustainability evaluation by a Pugh matrix also corroborated the sustainability of the long-term PCCI combustion with the PO addition. Compared to all of the fuel blends, CV10+P30%-CR19 placed the highest on the overall SPI (SPI = 1.28) for the same optimised CV10+P30%-CR19 case. Its analysis on weighted multi-criteria consideration of performance, emissions, renewability, carbon footprint, availability, engine compatibility, and safety. CV10–pine blends were always better than CV20–pine blends due to better atomisation and a more appropriate oxygen content. Neat diesel, on the other hand, received the lowest score in sustainability. The sustainability assessment showed that, alongside the advancement of efficiency and reduction of particulate matter, changing PO for other oil could decrease the carbon footprint while enhancing energy independence. The only tolerable NOx sacrifice.

The experiment–RSM–ML–TQM–Pugh matrix integrated framework provides an example of a technical-feasibility, eco-friendliness, and system sustainability-based optimised fuel-benefit integration pathway towards low-carbon CI engines where CV10 pilot fuel used with PO run at high compression ratios in PCCI is potentially feasible in practice. These findings provide strong support for the use of PO as a renewable fuel for advanced PCCI engines. In addition, future works should focus on life-cycle carbon assessment and durability analyses, in combination with the previously discussed NOx minimisation via either EGR or injection timing control, to establish the practical relevance of this new combustion concept.

Abbreviations

Abbreviation	Full Form
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
BTE	Brake Thermal Efficiency
BSFC	Brake Specific Fuel Consumption
COV	Coefficient of Variation
CO	Carbon Monoxide
CO_2	Carbon Dioxide
CR	Compression Ratio
CV10	10% Microalgae Biodiesel + 90% Diesel (Pilot Fuel)
CV20	20% Microalgae Biodiesel + 80% Diesel (Pilot Fuel)
D100	Neat Diesel
DMAIC	Define–Measure–Analyse–Improve–Control
DNN	Deep Neural Network
EGR	Exhaust Gas Recirculation
ELM	Extreme Learning Machine
GBM	Gradient Boosting Machine
GUI	Graphical User Interface
GPR	Gaussian Process Regression
HC	Unburned Hydrocarbons
HSU	Hartridge Smoke Unit
IC	Internal Combustion
KPI	Key Performance Indicator
LHV	Lower Heating Value
ML	Machine Learning
NOx	Oxides of Nitrogen
PO10%	10% Pine Oil
PO20%	20% Pine Oil
PO30%	30% Pine Oil
PCCI	Premixed Charge Compression Ignition
Q–Q Plot	Quantile–Quantile Plot
R^2	Coefficient of Determination
RSM	Response Surface Methodology
SHAP	SHapley Additive exPlanations
SPI	Sustainability Performance Index
SVM	Support Vector Machine
SVR	Support Vector Regression
TPI	TQM Performance Index
TQM	Total Quality Management
VCR	Variable Compression Ratio
XGBoost	Extreme Gradient Boosting

Conflict of Interest

The authors declare that they have **no known competing financial interests or personal relationships** that could have appeared to influence the work reported in this paper.

Funding Statement
The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work under Grant No. RGP2/306/44.

Acknowledgment: The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work under Grant No. RGP2/306/44.

Data availability: The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Author Contributions
Awadh: Conceptualization, RSM and statistical analysis, Machine learning modeling, Validation, Data curation, Writing – original draft.
Michael: Methodology, Experimental investigation, Writing – review & editing.

Reference

1. Jayakar, J., Gunasekar, N. & Elumalai, P. V. Combustion and emission analysis of hydrogen-microalgae biodiesel dual-fuel CI engine with cold EGR. *J. Therm. Anal. Calorim.* https://doi.org/10.1007/s10973-025-14879-1 (2025) doi:10.1007/s10973-025-14879-1.
2. Gurusamy, M. & Subramanian, B. Study of PCCI engine operating on pine oil diesel blend (P50) with benzyl alcohol and diethyl ether. *Fuel* **335**, 127121 (2023).
3. Allmägi, R., Ilves, R. & Olt, J. COMPREHENSIVE REVIEW OF INNOVATION IN PISTON ENGINE AND LOW TEMPERATURE COMBUSTION TECHNOLOGIES. *Transport* **39**, 86–113 (2024).
4. Bharadwaz, Y. D. & Kumari, A. S. PCCI combustion of low-carbon alternative fuels: a review. *J. Therm. Anal. Calorim.* **148**, 5179–5207 (2023).
5. Cao, D. N., Hoang, A. T., Luu, H. Q., Bui, V. G. & Tran, T. T. H. Effects of injection pressure on the NOx and PM emission control of diesel engine: A review under the aspect of PCCI combustion condition. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* **46**, 7414–7431 (2024).
6. Rameez, P. V. & Mohamed Ibrahim, M. A comprehensive review on the utilization of hydrogen in low temperature combustion

strategies: Combustion, performance and emission attributes. *Journal of the Energy Institute* **113**, 101511 (2024).

7. Le, T. T. *et al.* Partially-charged advanced low-temperature combustion in diesel engine: Progress and prospects. *Alexandria Engineering Journal* **129**, 373–440 (2025).
8. Hua, Y. Ethers and esters as alternative fuels for internal combustion engine: A review. *International Journal of Engine Research* **24**, 178–216 (2023).
9. Zhou, J., Chen, L., Zhang, R. & Zhao, W. Study on Oxidation Activity of Hydrogenated Biodiesel–Ethanol–Diesel Blends. *Processes* **12**, 462 (2024).
10. Muniyappan, S. & Krishnaiah, R. Investigation on PCCI combustion with cetane improver as secondary fuel for hybrid engine application. *Results in Engineering* **28**, 107372 (2025).
11. Rahman Adib, A., Mizanur Rahman, Md., Hassan, T., Ahmed, M. & Al Rifat, A. Novel biofuel blends for diesel engines: Optimizing engine performance and emissions with C. cohnii microalgae biodiesel and algae-derived renewable diesel blends. *Energy Conversion and Management: X* **23**, 100688 (2024).
12. Ihsan Shahid, M. *et al.* Hydrogen production techniques and use of hydrogen in internal combustion engine: A comprehensive review. *Fuel* **378**, 132769 (2024).
13. Zhu, J. *et al.* Sooting tendencies of terpenes and hydrogenated terpenes as sustainable transportation biofuels. *Proceedings of the Combustion Institute* **39**, 877–887 (2023).
14. CHIVU, R. M. The use of turpentine as additive for diesel oil. A review. *Journal of Thermal Engineering* 880–895 (2025) doi:10.14744/thermal.0000952.
15. Neupane, D. Biofuels from Renewable Sources, a Potential Option for Biodiesel Production. *Bioengineering* **10**, 29 (2022).
16. Rahman Adib, A., Mizanur Rahman, Md., Hassan, T., Ahmed, M. & Al Rifat, A. Novel biofuel blends for diesel engines: Optimizing engine performance and emissions with C. cohnii microalgae biodiesel and algae-derived renewable diesel blends. *Energy Conversion and Management: X* **23**, 100688 (2024).
17. Jhanani, G. K., Salmen, S. H., Al Obaid, S. & Mathimani, T. Performance and emission analysis of Scenedesmus dimorphus-based biodiesel with hydrogen as fuel in an unmodified compression ignition engine. *Int. J. Hydrogen Energy* **99**, 1132–1138 (2025).

18. Sindhu, R., Prabhat, S. T., Hiep, B. T., Chinnathambi, A. & Alharbi, S. A. Experimental assessment of cork based Botryococcus braunii microalgae blends and hydrogen in modified multicylinder diesel engine. *Fuel* **359**, 130468 (2024).

19. Veza, I., Spraggon, M., Fattah, I. M. R. & Idris, M. Response surface methodology (RSM) for optimizing engine performance and emissions fueled with biofuel: Review of RSM for sustainability energy transition. *Results in Engineering* **18**, 101213 (2023).

20. Lestari, W. D. *et al.* Optimization of 3D printed parameters for socket prosthetic manufacturing using the taguchi method and response surface methodology. *Results in Engineering* **21**, 101847 (2024).

21. Passerine, B. F. G. & Breitkreitz, M. C. Important Aspects of the Design of Experiments and Data Treatment in the Analytical Quality by Design Framework for Chromatographic Method Development. *Molecules* **29**, 6057 (2024).

22. Modi, V. *et al.* Machine learning and response surface optimization to enhance diesel engine performance using milk scum biodiesel with alumina nanoparticles. *Sci. Rep.* **15**, 34243 (2025).

23. Lionus Leo, G. M., Jayabal, R., Kathapillai, A. & Sekar, S. Performance and emissions optimization of a dual-fuel diesel engine powered by cashew nut shell oil biodiesel/hydrogen gas using response surface methodology. *Fuel* **384**, 133960 (2025).

24. Pan, X., Guan, W., Gu, J., Wang, X. & Zhao, H. Optimization of the low-load performance and emission characteristics for a heavy-duty diesel engine fueled with diesel/methanol by RSM-NSWOA. *Renew. Energy* **245**, 122819 (2025).

25. Li, J., Wang, H. & Dong, Q. Hybrid machine learning-based modeling of engine behavior using third-generation biodiesel: validation and robustness with SHAP explainability, bootstrapping, and sensitivity analysis. *Appl. Therm. Eng.* **281**, 128502 (2025).

26. Kamarulzaman, M. K. & Abdullah, A. Multi-objective optimization of diesel engine performances and exhaust emissions characteristics of *Hermetia illucens* larvae oil-diesel fuel blends using response surface methodology. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* **47**, 2952–2965 (2025).

27. Eskandari, H., Saadatmand, H., Ramzan, M. & Mousapour, M. Innovative framework for accurate and transparent forecasting of energy consumption: A fusion of feature selection and interpretable machine learning. *Appl. Energy* **366**, 123314 (2024).

28. Li, J., Wang, H. & Dong, Q. Hybrid machine learning-based modeling of engine behavior using third-generation biodiesel: validation and robustness with SHAP explainability, bootstrapping, and sensitivity analysis. *Appl. Therm. Eng.* **281**, 128502 (2025).

29. Alam, G. M. I. *et al.* Deep learning model based prediction of vehicle CO2 emissions with eXplainable AI integration for sustainable environment. *Sci. Rep.* **15**, 3655 (2025).

30. Houdou, A. *et al.* Interpretable Machine Learning Approaches for Forecasting and Predicting Air Pollution: A Systematic Review. *Aerosol Air Qual. Res.* **24**, 230151 (2024).

31. K. R, Latha. Rao. & Jagwani, S. Interpretable and Robust Machine Learning Approaches for Electric Vehicle Range Prediction Using SHAP Values. in *2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)* 1–6 (IEEE, 2025). doi:10.1109/IACIS65746.2025.11210912.

32. Chen, Y. *et al.* Machine learning-based design of target property-oriented fuels using explainable artificial intelligence. *Energy* **300**, 131583 (2024).

33. Yuan, D., Tang, L., Yang, X., Xu, F. & Liu, K. Explainable Machine Learning Prediction of Vehicle CO2 Emissions for Sustainable Energy and Transport. *Energies (Basel).* **18**, 5408 (2025).

34. Li, J., Wang, H. & Dong, Q. Hybrid machine learning-based modeling of engine behavior using third-generation biodiesel: validation and robustness with SHAP explainability, bootstrapping, and sensitivity analysis. *Appl. Therm. Eng.* **281**, 128502 (2025).

35. Žvirblis, T., Čižiūnienė, K. & Matijošius, J. Application of Machine Learning for Fuel Consumption and Emission Prediction in a Marine Diesel Engine Using Diesel and Waste Cooking Oil. *J. Mar. Sci. Eng.* **13**, 1328 (2025).

36. Psarommatis, F. & Azamfirei, V. Zero Defect Manufacturing: A complete guide for advanced and sustainable quality management. *J. Manuf. Syst.* **77**, 764–779 (2024).

37. Pawanr, S., Sarmah, P. & Gupta, K. TQM implementation for enhancement of product and service quality: a review on developments and latest trends. *Journal of Applied Research in Technology & Engineering* **6**, 63–72 (2025).

38. Waseem, M. & Yusoff, Y. M. The effect of total quality management practices on supply chain performance in the automobile industry. *Multidisciplinary Science Journal* **7**, 2025077 (2024).

39. Maganga, D. P. & Taifa, I. W. R. Quality 4.0 conceptualisation: an emerging quality management concept for manufacturing industries. *The TQM Journal* **35**, 389–413 (2023).

40. Antwi, S. & Darkwa Fentim, B. Total Quality Management and Organizational Performance: A Literature Review. *SSRN Electronic Journal* https://doi.org/10.2139/ssrn.4230846 (2021) doi:10.2139/ssrn.4230846.

41. DeHart, S. P. The Six Sigma Handbook, 4th edition. *Journal of Quality Technology* **47**, (2015).

42. Jensen, W. A. Response Surface Methodology: Process and Product Optimization Using Designed Experiments 4th edition. *Journal of Quality Technology* **49**, (2017).

43. Sharma, P. *et al.* Experimental investigations on efficiency and instability of combustion process in a diesel engine fueled with ternary blends of hydrogen peroxide additive/biodiesel/diesel. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* **44**, 5929–5950 (2022).

44. Sunil Kumar, K. *et al.* Performance, Combustion, and Emission analysis of diesel engine fuelled with pyrolysis oil blends and n-propyl alcohol-RSM optimization and ML modelling. *J. Clean. Prod.* **434**, 140354 (2024).