Results in Engineering 30 (2026) 110203

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

Multi-objective prediction and optimization of alcohol–gasoline SI engines using a hybrid gradient boosting and evolutionary algorithm framework

Muhammad Usman [a,*], Syed Muhammad Mehdi Ali Rizvi [a], Muhammad Sher Ali [a], Fahid Riaz [b], Muhammad Saeed [c], Essam M. Abo-Zahhad [d]

[a] Department of Mechanical Engineering, University of Engineering and Technology, Lahore, Pakistan
[b] College of Engineering, Abu Dhabi University, Abu Dhabi, , United Arab Emirates
[c] Abu Dhabi Maritime Academy, Abu Dhabi P.O. Box 54477, United Arab Emirates
[d] Department of Mechanical and Industrial Engineering, Liwa University, Abu Dhabi, 41009, United Arab Emirates

ARTICLE INFO

Keywords:
Biofuels
RSM
NSGA-III
Gradient boosting
Engine characteristics
Optimization
machine learning

ABSTRACT

Alcohol-blended fuels in spark-ignition (SI) engines offer an efficient way to enhance engine efficiency and reduce emissions. This study integrates experimental engine testing with a three-level data-driven model; Gradient Boosting to perform multi-output predictions, and Response Surface Methodology (RSM) and NSGA-III to determine operating points that offer a balanced outcome between performance and emissions in 1-propanol-gasoline blends. The multi-channel tests of an SI engine of one cylinder (1700–3800 rpm; two load settings) produced eight simultaneous responses: torque, brake power, BSFC, BTE, CO, CO_2, HC, and NO_x. The trained model was very accurate on most of the targets (e.g., BTE test R^2: 0.95; BP test R^2: 0.99). Both RSM and NSGA-III converged near 18% propanol and 2900–2950 rpm, with comparable Pareto-optimal performance. The 18% propanol mixture gave significant boosts as compared to pure gasoline at the same operating conditions: brake power was up by 31%, torque by 29%, CO and HC emissions were both reduced by 37%, and BSFC had risen by 19%. In conjunction with GB-RSM-NSGA-III methods, there exists a clear, repeatable approach to optimizing multiple objectives in alcohol blends for SI engines to promote clean-burning modes to fulfill SDG Goal 7 and SDG Goal 13.

1. Introduction

Internal combustion engines (ICEs) have served as the primary powertrain for mobile applications since the 1890s [1], yet they are also recognized as major contributors to CO_2 emissions in the transportation sector [2,3]. As global efforts toward carbon neutrality intensify, ICE vehicles face mounting pressure to comply with emission standards [4,5]. Winkler et al. [6] investigated the effect of vehicle emission regulations on air quality in the U.S., EU and China. The research methods involved analyzing historical emissions trends for criteria pollutants including PM, NO_x, CO and HCs from light-duty and heavy-duty vehicles. They concluded that stringent emission standards have reduced criteria pollutants by (80–95) % since the 1990s, significantly improving air quality. The ICEs should be modified to compete with electrified powertrains, such as battery electric vehicles. Berger and Jorgensen [7] investigated the carbon dioxide emissions from electric vehicles (EVs) compared to internal combustion vehicles (ICVs) and hybrid electric vehicles. The research methods included thermodynamic calculations and data analysis from the U.S. Environmental Protection Agency (EPA) and International Energy Agency (IEA). They concluded that emissions were relatively higher in case of ICVs in comparison to EVs. In response, emerging ICE technologies are increasingly focused on enhancing fuel efficiency, minimizing environmental impact, and cleaner-burning fuels to ensure continued relevance in a low-carbon future [8,9]. One viable pathway toward these goals is the partial replacement of gasoline with higher alcohols such as propanol, which offers promising combustion characteristics including high oxygen content, favorable volatility and a capacity to reduce regulated emissions [10]. Qian et al. [11] studied the effect of n-propanol as a partially fuel on combustion and emission characteristics in a direct injection spark ignition (SI) engine. The research methods included testing n-propanol/gasoline blends with 10%, 30% and 50% n-propanol by volume, maintaining RON at 95. They found that higher n-propanol ratios reduced CO and unregulated emissions and increased thermal efficiency. Kaisan et al. [12] investigated the effects of propanol and camphor blended with gasoline on the

* Corresponding author.
E-mail address: muhammadusman@uet.edu.pk (M. Usman).

https://doi.org/10.1016/j.rineng.2026.110203
Received 29 December 2025; Received in revised form 17 March 2026; Accepted 20 March 2026
Available online 20 March 2026

Nomenclature

Parameters	Abbreviation
BP	Brake Power
BSFC	Brake Specific Fuel Consumption
EPA	Environmental Protection Agency
EVs	Electric vehicles
ICEs	Internal combustion engines
IEA	International Energy Agency
ICVs	internal combustion vehicles
GB	Gradient Boosting
RSM	Response Surface Methodology
NSGA-III	Non-dominated Sorting Genetic Algorithm III
BTE	Brake Thermal Efficiency
SDG	Sustainable Development Goals
SI	Spark ignition

performance and emissions of a spark ignition engine. The research methods included testing various fuel blends in a four-stroke single--cylinder engine and analyzing exhaust emissions. They found out that blends with higher propanol content improved brake power (BP) and reduced HC and NO_x emissions. However, the integration of propanol into existing engine systems introduces new complexities. The combustion behavior of alcohol-gasoline blends is highly sensitive to operating condition, including engine load, fuel-air mixture, and ignition timing [13,14]. Moreover, the simultaneous optimization of multiple performance indicators like torque, brake power (BP), brake-specific fuel consumption (BSFC) and regulated emissions can be quite challenging. These factors are often connected in complex ways and changes in one can affect the others. Traditional physics-based models are useful for understanding basic engine behavior but often lack the flexibility and scalability needed for real time prediction and optimization in such multi-variable systems [15–17]. To address these challenges, data-driven modeling approaches have gained importance. These methods use tools like machine learning, statistics and optimization techniques to build models that can accurately predict engine behavior a wide range of operating conditions. Among these, methods like Gradient Boosting (GB) and Artificial Neural Network (ANN) are especially good at handling complex relationships and predicting several outputs at once [18–20]. When combined with statistical tools like RSM, these models offer a powerful framework for both prediction and optimization [21–23]. Aydin et al. [24] investigated the effect of biodiesel-diesel blends on the performance and emissions of a single-cylinder diesel engine. The research methods included experimental testing, artificial neural network (ANN) modeling and response surface methodology optimization. They concluded the approach effectively predicted engine parameters with high accuracy ($R2 > 0.89$), and optimal conditions were 32% biodiesel, 816 W load, and 470 bar injection pressure. Usman et al. [25] studied the effect of ethanol-gasoline blends on the performance and emissions of a spark ignition internal combustion engine. The research methods involved an artificial neural network (ANN) to predict engine responses and response surface methodology (RSM) to optimize performance while using various ethanol concentrations, engine loads, and speeds. They concluded that an 18.5% ethanol blend at 3700 RPM and 100% load significantly improved brake power and thermal efficiency by 19.9% and 29.8% respectively, while reducing CO and HC emissions by 28.1% and 40.6%. Moreover, integrating meta-heuristic evolutionary algorithms, such as the Non-dominated Sorting Genetic Algorithm III (NSGA-III), with statistical methods can play a key role in optimizing the engine's performance. Yang et al. [33] investigated the effect of a hybrid multi-objective optimization method on exhaust heat-management strategies for diesel engines at low loads. The research methods included RSM and NSGA-III with experimental validation using a Box-Bhenken design. They concluded that the optimized strategy significantly improved NOx conversion efficiency from 76.61% to 98.67%. Recent advancements in engine optimization have also explored the integration of nano-additives and specialized coatings to transcend traditional efficiency limits. For instance, studies on biodiesel engines have demonstrated that nanoparticles can act as oxygen-promoting catalysts, while thermal barrier coatings (TBC) reduce heat loss to the coolant, both of which have been modeled using machine learning to predict performance gains [26]. Despite advancements in alcohol–gasoline SI engine research, most studies rely on single predictive models (e.g., ANN) and single-objective optimization, limiting robustness and failing to capture real-world trade-offs between performance and emissions. Ensemble learning combined with evolutionary multi-objective optimization is not a well-explored field. Precisely, traditional ANN-based engine models are generally weak in generalizing on small experimental data and are highly subject to large variances. Moreover, single-objective or weighted desirability functions that are common in engine studies do not reflect the complicated, non-linear trade-offs that characterize many-objective engine systems. This paper will fill these gaps by using a Gradient Boosting ensemble to enhance predictive generalization by correcting sequence of errors and use NSGA-III to preserve diversity on a 8-objective Pareto front, which is highly efficient and mathematically rigorous than classical weighted methods. To address this gap, this study proposes a hybrid Gradient Boosting–based predictive model coupled with evolutionary multi-objective optimization, enabling improved generalization and simultaneous optimization of conflicting engine performance and emission parameters.

In this study, a validated data-driven approach for modeling prediction and simultaneous optimization of the engine's power output, thermal efficiency, and emissions for a gasoline-propanol spark ignition engine is provided. GB regression is integrated with RSM and NSGA-III to map trade-offs across speed, load, and blend ratio and to pinpoint robust operating points. The study also benchmarks the accuracy of traditional machine-learning models and demonstrates how genetic algorithms, along with RSM, can be used to identify optimal conditions (Table 1).

2. Materials and methodology

2.1. Experimental data collection

A test study was conducted in which an air-cooled spark-ignition engine with an overhead valve setup was used. A 7 inch dynamometer Dynometer was connected to the engine along with a water-brake load application system. Torque, BP, and BSFC were experimentally calculated and BTE was mathematically determined using BP and lower heating value of fuel. A 500 mL measuring cylinder (graduated to 0.5 mL precision) was used to measure instantaneous fuel flow. Emissions were measured with an EMS-5002 exhaust gas analyzer. Engine speeds were varied from 1700 to 3800 RPM with 300 RPM increments, at both 50% (10 psi) and 100% (20 psi) loading conditions. Gasoline was blended with different proportions of propanol immediately just before testing to prevent moisture absorption. These values ranged from 0% propanol-gasoline blend (pure gasoline) to 18% propanol-gasoline blend (18% n-propanol + 82% gasoline). The upper limit of propanol percentage (18%) for this research was chosen to remain within a low-to-moderate blending range. This ensured stable engine operation within the tested speed–load range and avoided hardware and/or calibration changes. The selected 0–18% interval also allowed sampling of separate blend ratios, allowing each blend to maintain its individual effects on the outputs while also capturing a broad range of combustion and emissions responses for surrogate training. Combining all of these, the experimental campaign comprised 112 unique operating points (2 load settings × 7 propanol–gasoline blend ratios × 8 engine-speed levels). Each test condition was repeated three times to ensure reproducibility, and

Table 1
Research novelty.

Author (s)	Year	Machine Learning Prediction	RSM	NSGA III	Fuel type (Propanol + Gasoline Blend)
Usman et al. [25]	2023	✔	✔	×	×
Godwin et al. [27]	2023	✔	×	×	×
Abdalla et al. [23]	2019	×	✔	×	×
Sahin et al. [28]	2025	✔	✔	×	×
Ahmad et al. [29]	2023	✔	✔	×	×
Singh et al. [30]	2019	×	✔	×	×
Onawumi et al. [31]	2019	×	✔	×	×
Badra et al. [32]	2022	×	✔	✔	×
Owoyele et al. [33]	2021	×	✔	✔	×
Sonawane et al. [34]	2023	✔	×	×	×
Yusri et al. [35]	2017	×	✔	×	×
Kapusuz et al. [36]	2015	✔	×	×	×
Thodda et al. [37]	2023	✔	✔	×	×
Moiz et al. [38]	2018	×	✔	✔	×
Yang et al. [39]	2025	×	✔	✔	×
Tirkey et al. [40]	2022	×	✔	×	×
Current study	2025	✔	✔	✔	✔

arithmetic mean of these values were recorded. DYNOMAX and Dynomite 2010 software were used for real-time data logging of speed, load, and fuel consumption. To ensure proper homogeneity, the blends were stirred continuously for 30 min under Standard Temperature and Pressure conditions (25 °C and 1 atm). Fig. 1 shows a schematic of the test rig

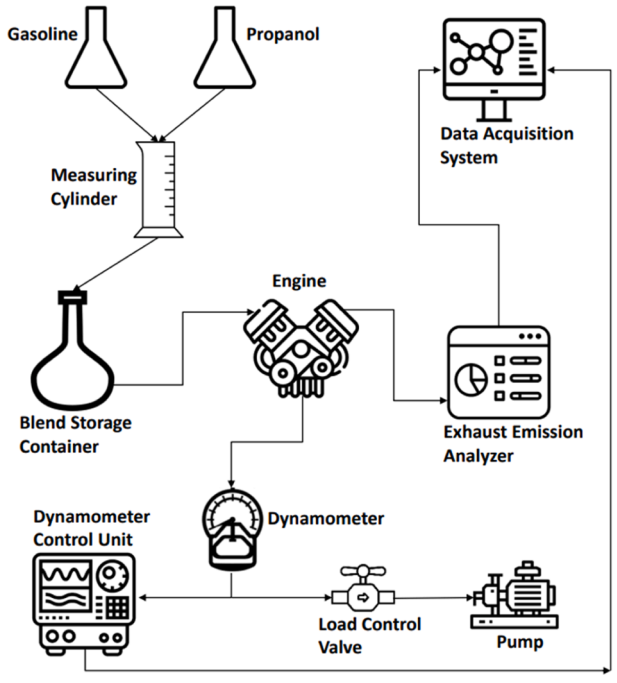

Fig. 1. Test rig layout and experimental configuration setup.

used for the experimentation.

Calculated metrics such as BP, BSFC, and BTE were derived using heating value of the fuels and standard thermodynamic equations. BP was calculated as:

$$BP = \frac{2\pi NT}{60 \times 1000} \tag{1}$$

Where N is engine speed in revolutions per minute (RPM), and (T) is the torque in Newton-meters. BTE was calculated using:

$$BTE = \frac{BP}{\dot{m_f} \times LHV} \tag{2}$$

where $(\dot{m_f})$ is the fuel mass flow rate and (LHV) is the lower heating value of the fuel. Eqs. (1) and (2) describe the fundamental thermodynamic computations for mechanical output and efficiency.

Similarly, BSFC was calculated using fuel mass flow rate and brake power. Eq. (3) describes the calculations of BSFC.

$$BSFC = \frac{\dot{m_f} \times 3600}{BP} \tag{3}$$

where $(\dot{m_f})$ is the fuel mass flow rate and (BP) is the brake power of the gasoline engine. Table 2 gives an overview about the properties of fuel that has been used in the experimentation of this study.

2.2. Data preprocessing and exploratory analysis

The raw dataset was first subjected to exploratory data analysis techniques, to understand the data patterns and highlight any abnormality beforehand. Histograms were made for each numeric variable to assess the parameter's distribution. Majority of the parameters were either normal, or had close to normal distribution (analyzed using Shapiro-Wilk distribution test). It is clear that the histogram follows a bell-shaped curve, and hence has no prominent outliers. Fig. 2, Fig. 3 and Fig. 4 show the data distribution through histograms of three of the output parameters.

However, BSFC, in particular, showed a non-normal distribution rampant with outliers. Fig. 3. shows the distribution of BSFC. The long right tail is visible in the histogram. Since parameters with such turbulent distributions often lead to inaccurate model training, BSFC was log-transformed for improved model performance.

The emission parameters had a close to normal distribution. Although, they did not pass the Shapiro-Wilk normalization tests, their histograms did not show any blatant outliers that warranted transformation of any kind.

Next, all input and output features were standardized using z-score normalization to ensure mean-centered data with unit variance.

In addition to the above techniques, a Pearson correlation heatmap was created to understand the relationship between all parameters, and to determine if two or more input features were highly correlated which has been shown in Fig. 5. Torque and CO_2 were most strongly correlated with BTE, which is feasible since higher CO_2 means better combustion, leading to a higher torque and consequently a higher BTE. Similarly, HC had a strong negative correlation with BTE. This is also in-line with existing understanding of combustion in ICEs, according to which Hydrocarbon emissions are caused by incomplete combustion.

Table 2
Properties of the fuel used in the study [13].

Property of Fuel	Test Methods	Gasoline	Propanol
Calorific Value (MJ/kg)	ASTM D240	45	33.6
Density (kg/m^3)	ASTM D4052	748	803
Oxygen Content (%)	ASTM D5622	0	27.6
RON	ASTM D2699	95	118

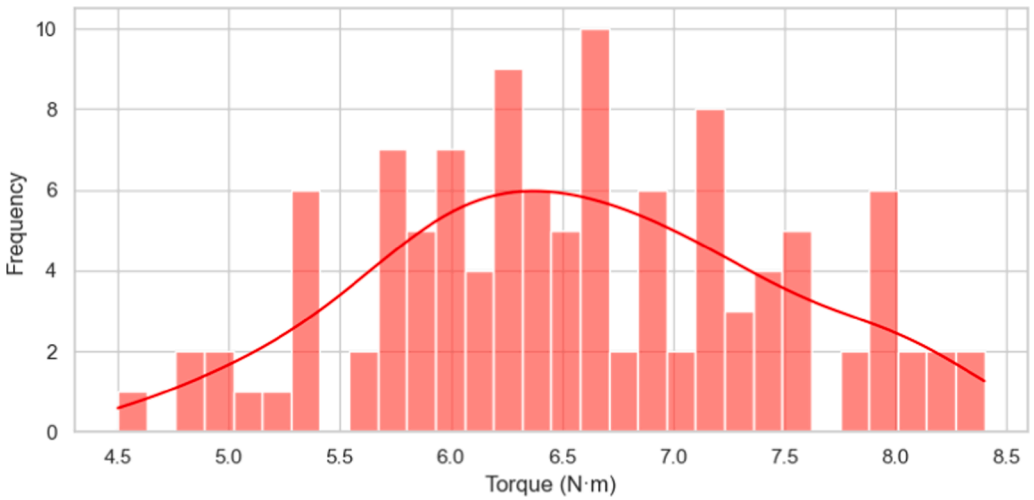

Fig. 2. Distribution of torque (N.m).

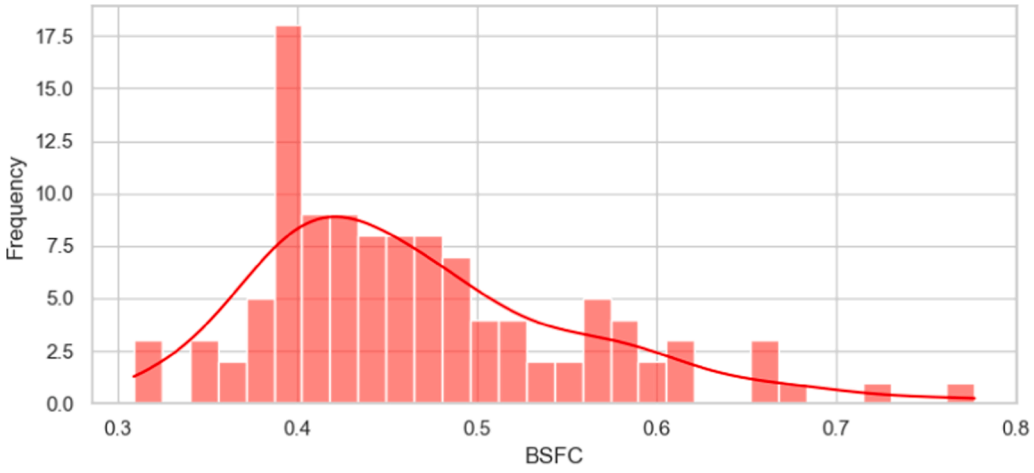

Fig. 3. Distribution of BSFC (kg/kW·h).

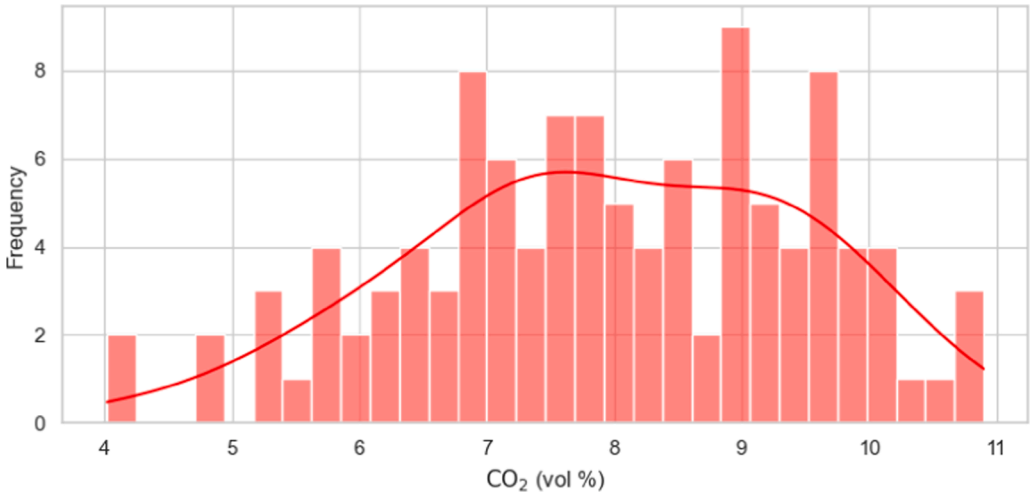

Fig. 4. Distribution of CO_2 (% Vol).

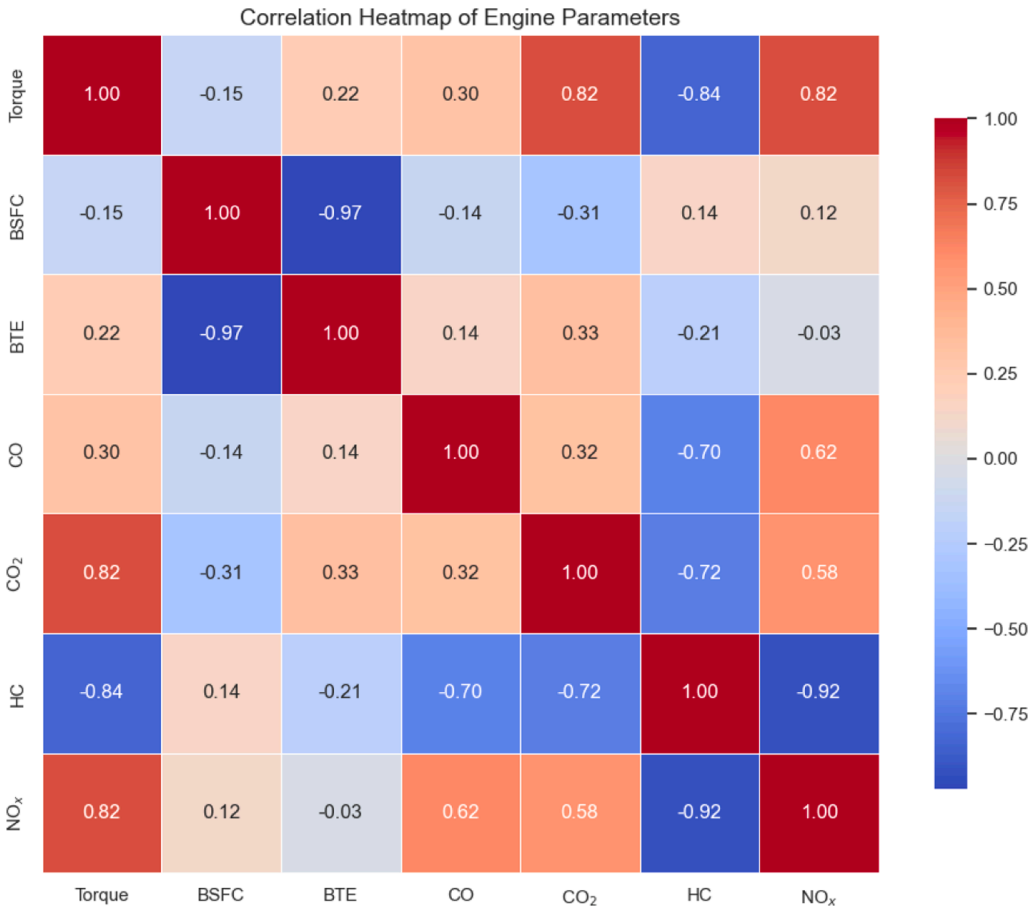

Fig. 5. Pearson correlation heat map.

2.3. Machine learning model development

2.3.1. Input and output features

The dataset included three input variables: Load (psi), Propanol (%), and Engine Speed (RPM). Eight output features were predicted which included Torque (N·m), Brake Power (kW), BSFC (kg/kW·h), Brake Thermal Efficiency (%), CO (% Vol), CO_2 (% Vol), HC (ppm), and NO_x (ppm).

2.3.2. Model selection

The next logical step was to model selection. Starting from the simplest linear regression, around 15 models were evaluated using LazyPredict library which provided an insight into what kind of model learned the underlying patterns governing the biofuel combustion data. LazyPredict is a lightweight AutoML-like Python library that helps in model selection by automatically testing various models without having to write boilerplate code for any. Using LazyRegressor, the library was

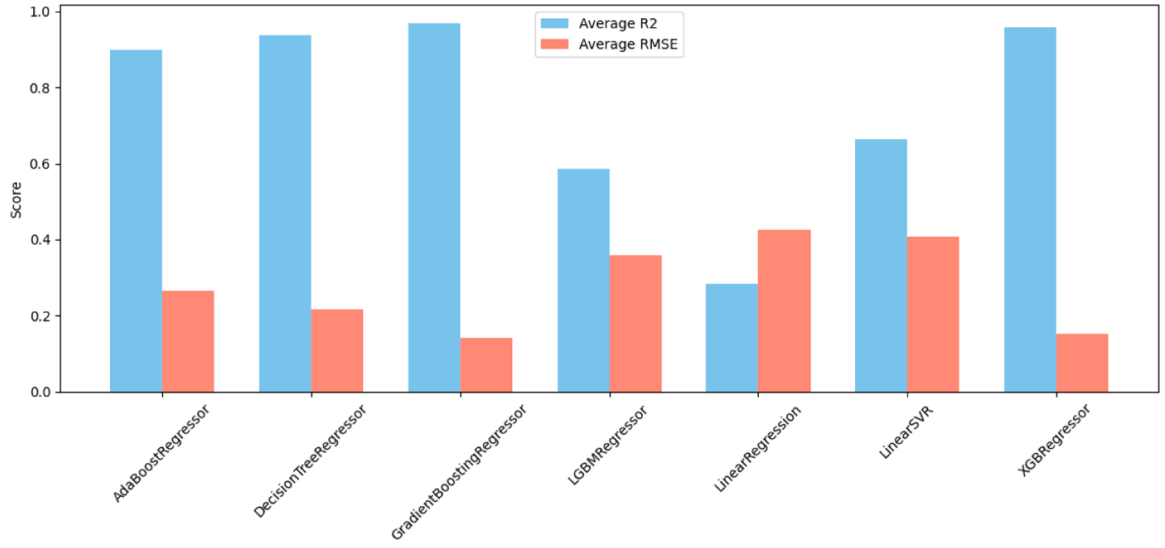

Fig. 6. Average R^2 and RMSE per model.

used to train many baseline models using their default configurations and a performance summary was generated. Each of the 15+ models were evaluated for each output individually. The top-performing model was shortlisted when they had average R2 scores for all outputs. The vertical bar chart in Fig. 6 provides a visualization of the comparison of seven important models that range from simple linear regressors to complicated boosting algorithms. LazyPredict framework helped to benchmark >15 algorithms for regression like Linear Regression, Decision Trees, Adaptive Boosting (AdaBoost), Extra Trees, eXtreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LGBM).

Gradient Boosting was selected since it showed the highest average predictive performance across the eight outputs during preliminary benchmarking. It also allowed the inclusion of a strong bias–variance control through hyperparameter tuning $(d_{max}, s_{split}, s_{leaf})$, among others. Gradient Boosting also achieved the highest average test R^2 (~0.96) and lowest average RMSE across all eight targets during the LazyPredict screening phase.

The other ML models were seen lagging behind this performance. The second-best ensemble model, XGBRegressor, had an R^2 0.8% lower than GB, and 4% lesser normalized RMSE. Linear regression and LinearSVR produced average R^2 values of 0.65–0.70, considerably lower than GB. Decision Trees and AdaBoost achieved intermediate performance (average $R^2 \approx 0.85$–0.89) but with higher inter-fold variance, leading to a higher chance of overfitting on the present dataset size. LGBMRegressor performed comparably to XGBoost overall but produced marginally higher RMSE for BSFC specifically, which was a deciding factor given the targeted Huber-loss robust head strategy adopted in this study.

2.3.3. Model training and tuning

A Gradient Boosting Regressor (GBR) wrapped in a MultiOutputRegressor was employed for multi-target regression. Hyperparameter tuning was performed using **Optuna**, a Bayesian optimization framework using the Tree-structured Parzen Estimator (TPE). The objective function minimized cross-validated Mean Squared Error (MSE) over 5 folds. To ensure reproducibility the components were controlled using fixed seeds (random_state = 42 for shuffling/splitting and cross-validation, TPESampler (seed = 42) for Optuna tuning, and seed = 42 for NSGA III).

2.3.4. Gradient boosting hyperparameter tuning

The finalized dataset (N = 112 averaged operating points) was randomly shuffled and split into 80% training and 20% held-out testing. Hyperparameter tuning was conducted using cross-validation on the training subset, and the final model was refit on the full training subset. Finally, it was evaluated on the unseen test set.

The gradient boosting model used in this study was optimized by tuning a set of hyperparameters $\theta = \{n_{estimators}, \eta, d_{max}, s_{split}, s_{leaf}\}$ using a 5-fold cross-validation routine. The objective of optimization was to maximize predictive performance by minimizing the average mean squared error (MSE) across folds.

After several iterations, the search space was designed to be under the following constraints: the number of estimators $n_{estimator}$ was selected from the integer range [50,500], while the learning rate $\eta \in [0.01, 0.3]$ controlled the contribution of each tree. Another parameter, tree depth d_{max} was varied in range 3–10. Node splitting constraints were defined by $s_{split} \in [2,20]$ for the minimum number of samples required to split an internal node in this tree-based gradient boosting model, and $s_{leaf} \in [1, 20]$ for the minimum number of samples required to form a leaf. All of these hyperparameters (along with their ranges) were chosen to make sure that the model does not overfit to the training data. This combination of hyperparameters enabled the optimizer to navigate a structured, high-dimensional search space and improve generalization for the multi-output regression employed in the study.

Optuna tuning resulted in the final Gradient Boosting configuration which was eventually used in this study:

$n_{estimator} = 479, \eta = 0.1351, d_{max} = 4, s_{split} = 6$, and $s_{leaf} = 4$.

For BSFC, a separately trained Gradient Boosting model using Huber loss was used to reduce the sensitivity of heavy-tailed BSFC.

2.3.5. Model evaluation

Model performance was quantitatively assessed using four mainstream error metrics: mean absolute error (MAE), mean squared error (MSE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2). For each target variable y_j, and corresponding predictions $\hat{y}_j$, the metrics were computed over the test set D_{test} of size N. Eqs. (4–7) outline these metrics formally:

$$MAE_j = \frac{1}{N}\sum_{i=1}^{N}\left|y_{ij} - \widehat{y_{ij}}\right| \tag{4}$$

$$MSE_j = \frac{1}{N}\sum_{i=1}^{N}\left(y_{ij} - \widehat{y_{ij}}\right)^2 \tag{5}$$

$$MAPE_j = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_{ij} - \widehat{y_{ij}}}{y_{ij}}\right| \tag{6}$$

$$R_j^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_{ij} - \widehat{y_{ij}}\right)^2}{\sum_{i=1}^{N}\left(y_{ij} - \underline{y_j}\right)^2} \tag{7}$$

These metrics determined the model's absolute error (MAE), squared error (MSE), percentage error (MAPE), and explanatory power (R^2) for each output variable (y_j). In this context, (y_j) denotes the average true value of the jth output over D_{ts}. MAE and MSE characterize the prediction error, with the latter disproportionately penalizing large deviations. We also employed MAPE to ensure interpretability across different scales, alongside the R^2 score to quantify the proportion of variance explained by the model.

2.4. Response surface modeling and optimization

The experimental data obtained from the engine tests mentioned were imported into Design-Expert Software for response surface methodology (RSM) modeling. We employed second-order polynomial regression (including interaction terms) to develop surrogate models for each of the 8 response variable. Following an evaluation of various optimization trajectories, the desirability function parameters were finalized. The specific goals and constraints utilized for the multi-objective optimization process are detailed in Table 3.

CO_2 (as a maximizing objective) was included as a proxy for more

Table 3
Desirability metrics and constraints.

Name	Goal	Lower Limit	Upper Limit	Lower Weight	Upper Weight
A: Load	is in range	10	20	1	1
B: Propanol	is in range	0	18	1	1
C: Engine Speed	is in range	1700	3800	1	1
Torque	maximize	4.5	8.4	1	1
BP	maximize	0.801105	3.18348	1	1
BSFC	minimize	0.309067	0.776917	1	1
BTE	maximize	10.456	25.8843	1	1
CO	minimize	0.95	9.4	1	1
CO_2	maximize	4.02	10.9	1	1
HC	minimize	35	290	1	1
NO_x	minimize	290	2000	1	1

Table 4
NSGA-III parameters.

Parameter	Value
Number of Objectives (M)	8
Partitions (H)	4
Reference Directions	330 (Das-Dennis method)
Population Size	330
Number of Generations	1200
Crossover Operator	SBX (probability 0.9, eta 15)
Mutation Operator	Polynomial (probability 0.333, eta 20)

complete combustion (consistent with its positive correlation with torque/BTE and expected inverse relation to CO/HC in incomplete-combustion scenarios); however, since CO_2 is a greenhouse gas, this objective should be seen as a combustion-completeness indicator rather than a signal of reducing climate impact.

2.5. NSGA-III analysis

Once the RSM was complete, the same optimization problem was solved with Non-dominated Sorting Genetic Algorithm III (NSGA-III). NSGA-III was selected since it is found useful in handling many-objective problems (eight objectives in this study) through structured reference directions and promoting diversity in the Pareto front. The optimization problem was formulated with three decision variables bounded within the training data range to avoid extrapolation, the same range that was used in RSM: load (10, 20) psi, propanol (0, 18) %, and engine speed (1700, 3800) RPM. Eight conflicting objectives were used: minimization of BSFC, CO, HC, and NO_x; maximization of BTE, CO_2, torque, and brake power, similar to the constraints used for RSM. The NSGA-III algorithm was configured using the pymoo framework (version 0.6.0) to optimize the eight-objective problem. Reference directions were generated via the Das-Dennis method with 8 objectives (M) and 4 partitions (H), yielding 330 directions and a matching population size. The algorithm was seeded for reproducibility and terminated after a fixed number of generations and the key parameters for it are summarized in Table 4.

The parameter 'M' was used to set the objective space dimension, and 'H' determined the number of reference directions - together, both these parameters guided diversity across tradeoffs to create balanced solutions per generation.

The NSGA-III fitness evaluations were only done with the trained Gradient Boosting surrogate. The RSM surrogate models developed were not used within the NSGA-III optimization loop. The two methods are therefore an independent and complementary analysis of the same design space, and their convergence to similar optimal operating conditions is a mutual cross-validation for both approaches.

2.6. Statistical analysis using ANOVA and experimental validation

This research also performed analysis of variances (ANOVA) on each surrogate model using p-values together with predicted and adjusted coefficient-of-determination (R^2). The percentage contribution (PC%) of each term has been calculated based on the sum of squares in Design Expert Software. As per standard statistical procedures, significance was tested at the 95% confidence level.

The best settings that resulted from the RSM simulation were physically tested on a real engine. All eight outputs were measured - and the results were compared with the predicted values.

Fig. 7 shows a comprehensive overview of the overall methodology of this research study.

2.6.1. Measurement uncertainty, repeatability, and statistical limitations

All experimental measurements are prone to changes due to instrument uncertainty and variations in each experimental run. To reduce the effect of random error and to increase repeatability, each operating point was sampled thrice with the reported values being the arithmetic mean of the said runs. The measurement ranges, individual instrument accuracies and calculated relative uncertainty percentages for each parameter are mentioned in Table 5.

Engine performance data is dependent primarily on three things: The accuracy of the Dynomite dynamometer, the stability of the calibration of the EMS-5002, and the ambient conditions in the testing (experimental) cell. To validate the results, systematic uncertainty analysis has been performed by using the Holman error propagation method. In this method, independent as well as coupled variables are taken into account. For example, Brake power was calculated as a derived quantity from engine speed and torque (measured simultaneously), according to Eq. (1). The uncertainty in brake power was propagated from both measurements of speed and torque, with a combined uncertainty of ± 1.118 per cent.

Secondly, BSFC was computed as the ratio of measured fuel consumption to brake power. Since this output depends on measurements taken from three instruments simultaneously (volumetric gauge, dynamometer, and speed sensor.), its propagated uncertainty came out as ±1.5%.

The individual uncertainty values from all major instruments were combined to calculate the total uncertainty. These included the Hall-effect speed sensor, the dynamometer load cell, the volumetric fuel measurement system, and the NDIR and electrochemical sensors in the exhaust gas analyzer. The overall experimental uncertainty was determined by the Root-Sum-Square (RSS) method, as defined in Eqs. (8–12). This way the combined effect of all the measured parameters is captured. Since BSFC is a derived quantity dependent on fuel consumption and brake power, its uncertainty is implicitly captured through UFuel and UBP, and is therefore excluded from the RSS to avoid double-counting.

$$U_{Total} = \sqrt{(U_{Speed}{}^2) + (U_{Torque}{}^2) + (U_{Fuel}{}^2) + (U_{BP}{}^2) + (U_{CO}{}^2) + (U_{CO_2}{}^2) + (U_{HC}{}^2) + (U_{NO_X}{}^2)} \quad (8)$$

Substituting the corresponding uncertainty values from Table 5 in Eq. (9)

$$U_{Total} = \sqrt{(0.5^2) + (1.0^2) + (1.0^2) + (1.118^2) + (0.2^2) + (0.2^2) + (0.2^2) + (0.2^2)} \quad (9)$$

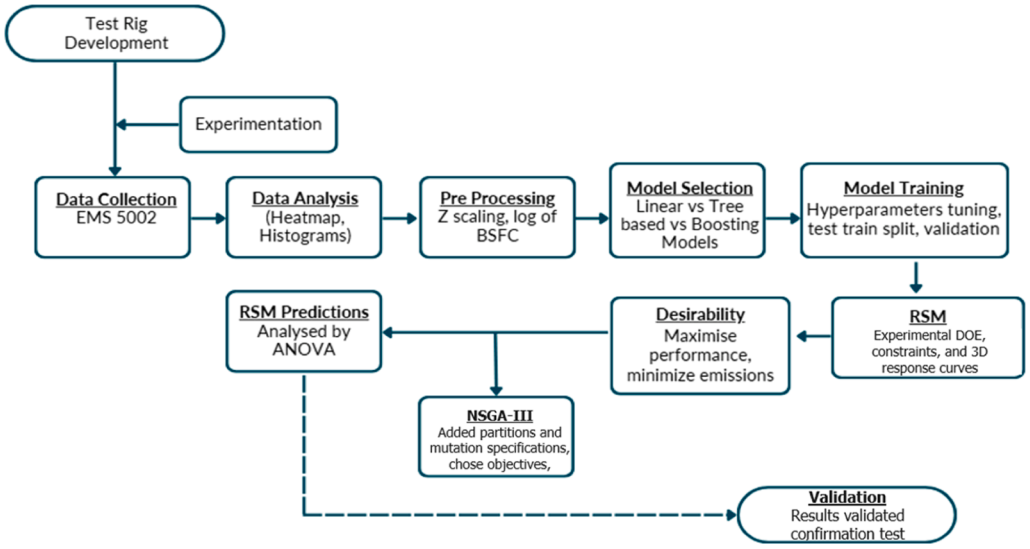

Fig. 7. Research workflow schematic.

Table 5
Uncertainty analysis.

Parameter	Measurable Range	Accuracy	Uncertainty (%)
Speed	0–8000 RPM	±1 RPM	±0.5
Torque	0–45 Nm	±0.1 Nm	±1.0
Fuel Consumption	-	±0.1 mL	±1.0
BSFC	-	±0.1 mL, ±0.1 Nm, ±1 RPM	±1.5
Brake Power	0–37.7 kW	±0.1 Nm, ±1 RPM	±1.118
CO	0–18.00% vol	±0.01% vol	±0.2
CO_2	0–18.00% vol	±0.01% vol	±0.2
HC	0–5000 ppm	±1 ppm	±0.2
NO_x	0–5000 ppm	±1 ppm	±0.2

Table 6
Performance metrics for engine model predictions on both the training and testing datasets.

Parameter	Set	MSE	MAE	R^2	MAPE
Torque	Training	0.0016	0.032	0.9979	0.005
	Test	0.016	0.0973	0.9776	0.0146
BP	Training	0.0001	0.0094	0.9996	0.0054
	Test	0.0028	0.0398	0.9936	0.0197
BSFC	Training	0.0003	0.01	0.9569	0.0216
	Test	0.0014	0.0263	0.801	0.0565
BTE	Training	0.0085	0.0722	0.9985	0.004
	Test	0.1674	0.3297	0.9538	0.0175
CO	Training	0.0029	0.0412	0.9993	0.0113
	Test	0.0515	0.1614	0.9892	0.0589
CO_2	Training	0.0055	0.0606	0.9978	0.008
	Test	0.042	0.1476	0.9737	0.0183
HC	Training	3.8102	1.5914	0.9989	0.0132
	Test	79.0085	6.7563	0.9672	0.0581
NO_x	Training	197.183	11.0493	0.9985	0.0136
	Test	2167.2	35.8625	0.9875	0.0356

Expanding the squared terms in the equation above gives us the final value.

$$U_{exp} = \sqrt{0.25 + 1.00 + 1.00 + 1.25 + 0.04 + 0.04 + 0.04 + 0.04} \quad (10)$$

Summing up the terms, we get:

$$U_{exp} = \sqrt{3.66} \quad (11)$$

Final computed uncertainty is seen as:

$$U_{exp} = 1.913\% \quad (12)$$

The total experimental uncertainty comes out as 1.913%. This value falls within the generally accepted tolerance limit of ±5% for ICE research.

Statistical significance was calculated with the help of p-values. The practical significance of each parameter was further assessed using the percentage contribution (PC%). This was defined as the ratio between the sum of squares of each factor and the corrected total sum of squares. The PC% metric (equivalent to an η^2 effect size) was used to quantify the proportion of response variation attributed to each input variable. This approach avoided over-interpretation of statistically significant terms that had minimal physical impact on the system.

3. Results and discussion

3.1. Machine learning performance

The Gradient Boosting model exhibited very good predictive accuracy on each of the 8 outputs. On the test set, R^2 for the output parameters were measured to be > 0.90 for most of the outputs, with strong performance indications for Brake Power ($R^2 = 0.9936$) and BTE ($R^2 = 0.9538$). Table 6 demonstrates that the model shows nearly-perfect metrics, indicating a good fit on the training data.

However, BSFC showed a relatively low R^2 score (0.801). BSFC is a ratio-derived quantity as evident from Eq. (3) - whose propagated measurement uncertainty (±1.5%, Table 5) is the highest of any parameter in this study. It also resulted in the production of the heavy-tailed distribution visible in Fig. 3. Small absolute errors in either the fuel flow numerator or the brake power denominator compound non-linearly, amplifying prediction error. The mitigation strategies employed (log-transformation) prior to training and a dedicated robust Gradient Boosting head with Huber loss - yielded a meaningful improvement. Quantitatively, the final model achieves MAE = 0.0263 kg/kWh, RMSE = 0.0423 kg/kWh, and MAPE = 5.65% on the unseen test set, placing the RMSE at approximately 9% of the full experimental BSFC range (0.31–0.78 kg/kWh). The learning curve in Fig. 9(b) shows cross-validation MSE converging toward training MSE as dataset size increases. This confirms that predictions are bounded by data density and that the model is not overfit.

To prevent overfitting, the hyperparameter tuning was done while steering the model towards strong regularization. Tree depth and leaf

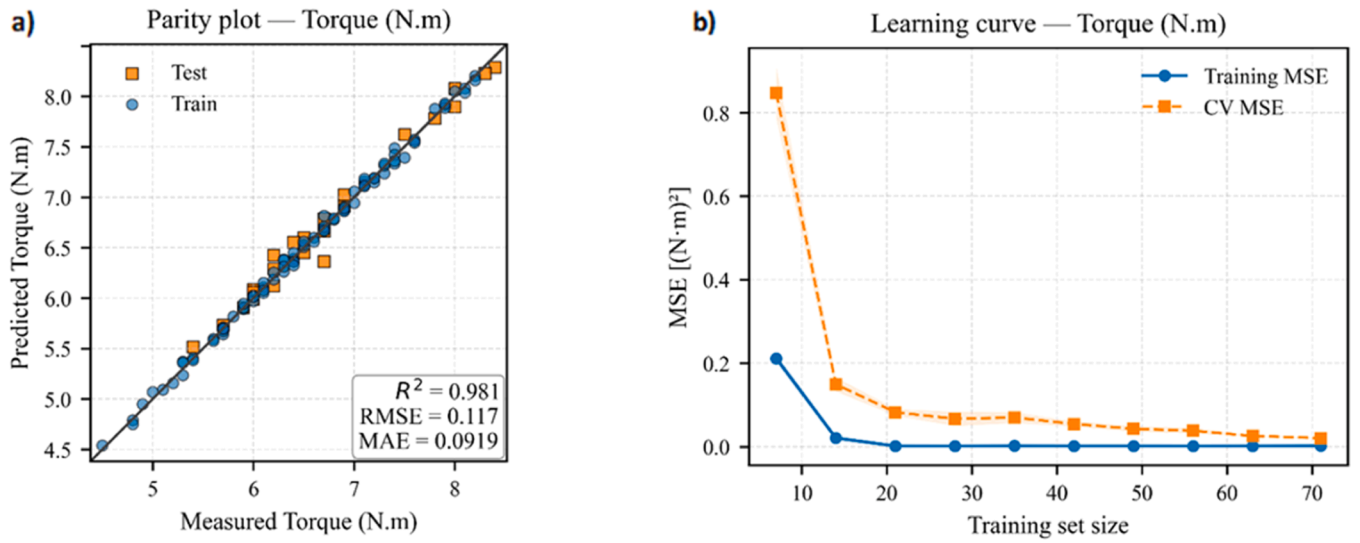

Fig. 8. a) Parity plot-torque (predicted versus actual); b) learning curves-torque.

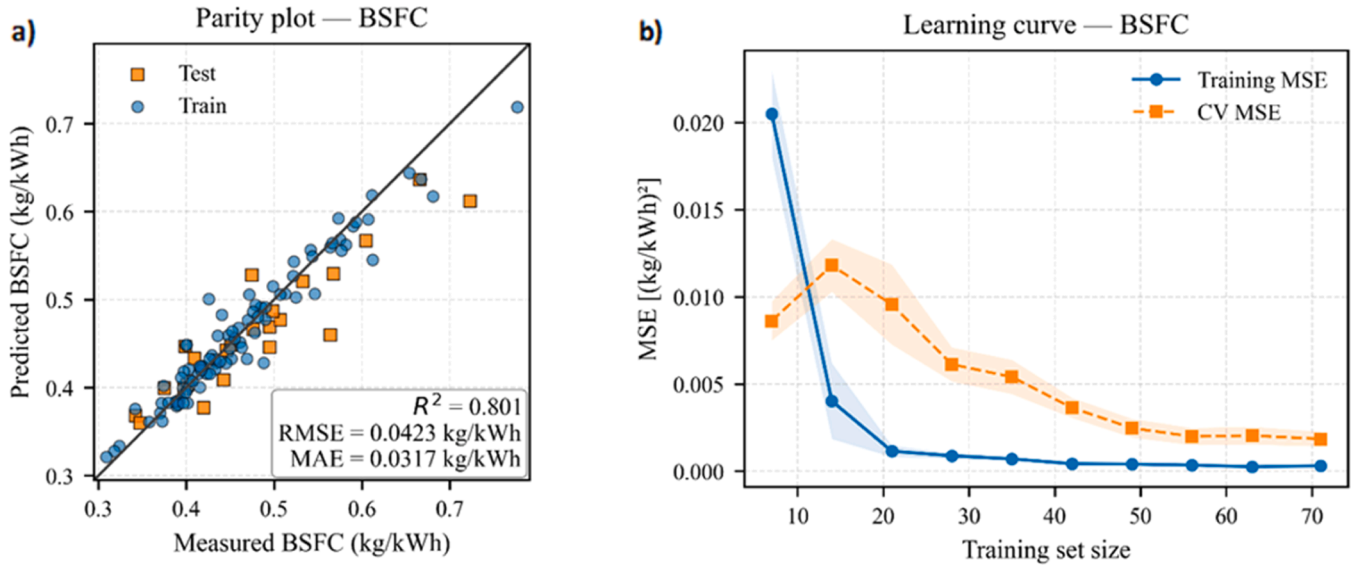

Fig. 9. a) Parity plot-BSFC (predicted versus actual); b) learning curves-BSFC.

splits were therefore constrained ($d_{max} = 4, s_{split} = 6, s_{leaf} = 4$) and stochastic subsampling was added to improve generalization. Hyperparameters were optimized using 5-fold cross-validation on the training dataset exclusively (the test set was withheld throughout all tuning iterations).

All final performance metrics were reported exclusively on an independent test set. The mean training R^2 across all eight outputs is 0.9934 and the mean test R^2 is 0.9561, yielding a mean generalization gap of 3.73%. For the four outputs most relevant to engine optimization (Torque, BP, BTE, CO), the train–test R^2 gap does not exceed 1.9 percentage points. Convergent learning curves were observed across all targets (Fig. 8b and Fig. 9b). This proves that the high training accuracy reflects accurate model predictive prowess and not overfitting.

Since the Gradient Boosting regressor is a deterministic model, it does not provide prediction intervals. Instead, we characterized the model's predictive uncertainty using the test-set RMSE values provided in Table 6. For the optimized solutions, the 95% confidence interval $CI_{95\%}$ is approximated as:

$$\hat{y} \pm 1.96 \times RMSE_{test} \quad (13)$$

To validate this approximation, bootstrap confidence intervals were constructed using $B = 200$ resamples of the training data; the resulting 95% half-widths were within approximately 1.5 × the RMSE-based estimate for all outputs, confirming that Eq. (13) provides a conservative (slightly over-wide) coverage bound. At the RSM-optimized operating point, the absolute CI half-widths turned out to be: Torque ±0.248 N·m, BP ±0.104 kW, BSFC ±0.073 kg/kWh, BTE ±0.802%, CO ±0.445 vol%, CO_2 ±0.402 vol%, HC ±17.43 ppm, and NOx ±91.24 ppm. These values are consistent with the instrument-level uncertainties reported in Table 5.

3.1.1. *Parity plots and learning curves*

Parity plots are scatter charts of predicted versus actual responses. These plots serve as a simple visual assessment of model accuracy. In these plots, points lying exactly on the 45° line indicated perfect predictions. While parity plots and R^2 captured overall fit, learning curves based on MSE revealed how the model's error evolved with increasing training data. By plotting training and cross validation MSE against training set size, underfitting (both curves high and converging) or overfitting (low training MSE but substantially higher validation MSE)

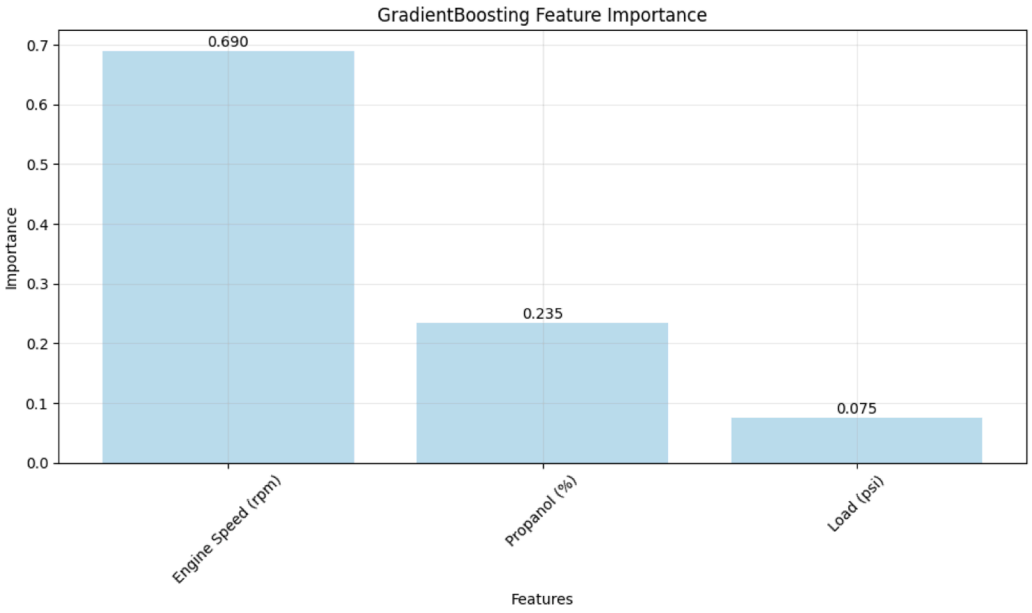

Fig. 10. Feature importance (input parameters).

can be diagnosed. The Figs. 8–9 show the parity plots and learning curves for two responses – torque and BSFC.

The parity plots for torque shown in Fig. 8(a) and BP shown in Fig. 9 (a) show similar trends; an excellent fit with a high R^2 show that the model has learned to accurately predict the results for these parameters. Their learning curves for torque and BP depicted in Fig. 8(b) and Fig. 9 (b) respectively show no signs of underfitting or overfitting.

The parity plots for BSFC as in Fig. 9(a) shows a moderate fit. Even though BSFC's plot does not show the high R^2 as shown by the other parameters, it still stands strong at 0.801. The learning curve of BSFC as shown in Fig. 9(b) shows the model converging at the end and the test MSE approaching 0. With additional data points, these numbers can be improved, and the model can be fit better than its current iteration.

Comprehensive plots for the remaining six outputs (BP, BTE, CO, CO_2, HC, and NO_x) display trends similar to Torque and are provided in the Supplementary Material (Figs. S1–S6).

3.1.2. *Feature importance*

The bar chart feature-importance analysis provides a coherent picture of the importance of all 3 input parameters in driving the model's prediction. From Fig. 10, it can be observed that engine speed is the dominant predictor, followed by propanol concentration, and finally by load. The feature importance for this engine-emissions study shows how much each of the 3 input features reduced the model's loss when it was used to split decision-tree nodes across all estimators.

3.2. *RSM desirability optimization*

The RSM stage of this research was formulated as a multi-response optimization under clearly defined input and output constraints. The decision variables and target responses have been summarized in Table 3. Load (10–20 psi), Propanol % (0–18%), and Engine Speed (1700 – 3800 RPM) were each constrained "in range" with equal weights and high importance. Simultaneously, outputs (responses) such as Torque, BP, BSFC, BTE, CO, CO_2, HC and NO_x were assigned goals of "maximize," "minimize," or "in range" according to engine performance and emission objectives. In line with the goals of this research, engine performance parameters were maximized, while emissions were minimized.

Model adequacy for each response was assessed through adjusted and predicted R^2 statistics which is shown in Table 7. Most RSM models showed good fit, with their R^2 above 0.95, indicating accurate surface modelling. The BSFC model showed a lower, but still above-par predictive capability with adjusted $R^2 = 0.9135$ and predicted $R^2 = 0.8366$. A total of 94 feasible solutions satisfying the constraints were generated by desirability-driven search. The top-ranked solution occurred at Load of 20 psi, Propanol percentage of 18%, and Engine Speed of 2930 RPM, with a composite desirability of 0.723. The corresponding predicted responses shown in Table 8 confirm that the algorithm identified a high-performance operating region for engine operation within the search space. Importantly, the optimization minimized emissions while maximizing fuel conversion efficiency and output power.

Table 7
Adjusted R^2 and Predicted R^2 for responses analyzed in RSM.

Response	Adjusted R^2	Predicted R^2
Torque	0.9527	0.9481
BP	0.9882	0.9870
BSFC	0.9135	0.8366
BTE	0.9647	0.9619
CO	0.9935	0.9927
CO_2	0.9390	0.9312
HC	0.9730	0.9690
NO_x	0.9823	0.9801

3.2.1. *RSM response curves*

All 8 responses (output parameters) were analyzed using RSM curves at low load of 10 psi and relatively high load of 20 psi, allowing us insights into their relationship with the input parameters. Fig. 11(a) and Fig. 11(b) show torque as a function of propanol percentage and engine speed at two fixed loads. Increasing propanol % generally elevates the surface, while torque rises with speed into a peak region and then moderates at the highest speeds, consistent with prior RSM trends.

Fig. 12(a) and Fig. 12(b) show the effects of changing propanol percentage, engine speed and load at brake power. BP showed a consistently linear relationship with propanol percentage and engine speed, with increasing both input parameters increase the BP. This trend remains consistent at both loads.

The RSM surfaces exhibit a saddle-like-shape - with a BSFC minimum around mid-speed as shown in Fig. 13(a) and Fig. 13(b). At low speeds, BSFC rises, especially at high propanol percentages, because charge cooling and poorer vaporization slow the burn and require more fuel per

Table 8
Best input and output parameters.

Load	Propanol	Engine Speed	Torque	BP	BSFC	BTE	CO	CO_2	HC	NO_x	Desirability
20	18	2930.233	8.01	2.457	0.37	22.177	3.6	9.89	82.039	1366.12	0.723 Selected

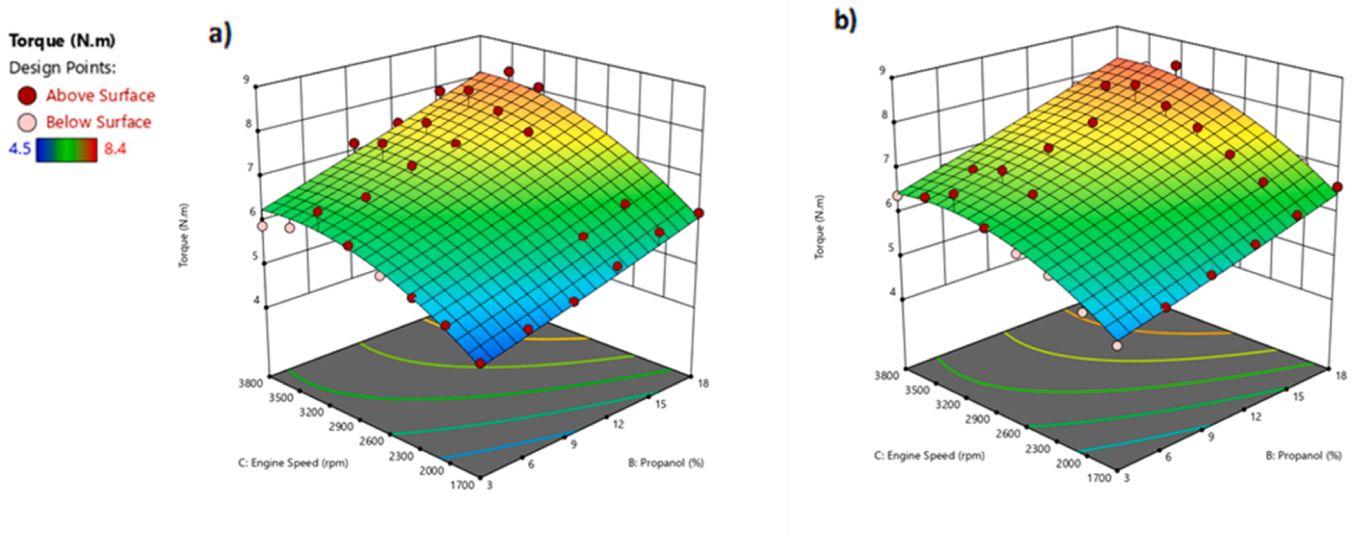

Fig. 11. a) Torque response curve at low load; b) Torque response curve at high load.

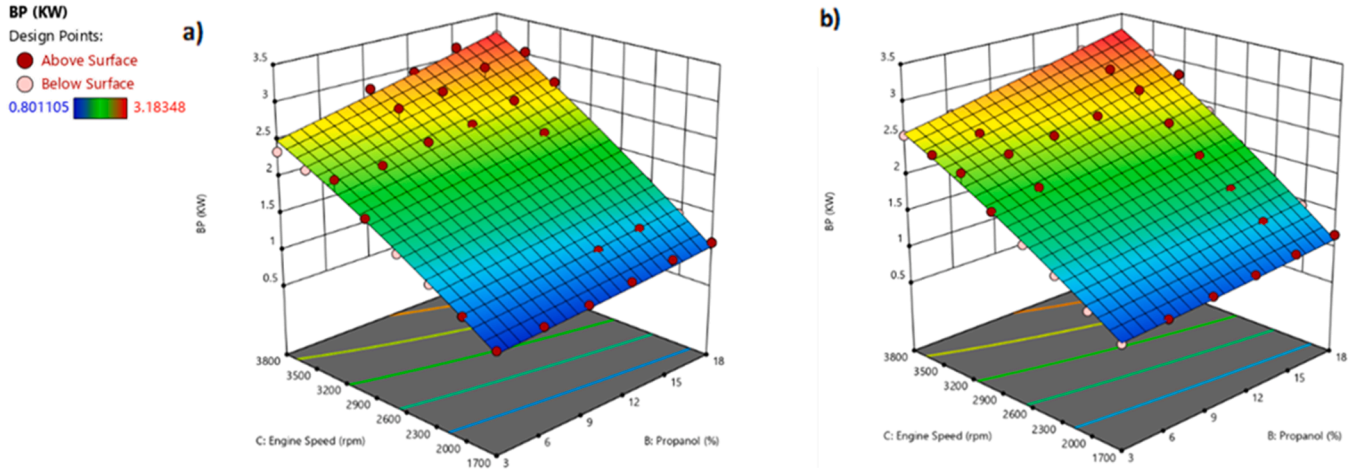

Fig. 12. a) BP response curve at low load; b) BP response curve at high load.

kWh. At very high speeds, rising mechanical/friction losses, as well as shorter combustion durations, again penalize BSFC. With higher load, the low-BSFC valley deepened and shifted toward higher propanol percentage, as the hotter in-cylinder environment offset propanol's latent-heat penalty [11,36] and the fuel-bound oxygen promoted more complete combustion; nevertheless, the lower heating value still limited gains at the extremes.

The BTE response curves in Fig. 14(a) and Fig. 14(b) reveal a relatively pronounced load dependence on the efficacy of propanol-gasoline blends, compared to other responses. At a low load, BTE is markedly suppressed, with peak efficiency (~22%) only attainable within a narrow operational band of high engine speed and minimal propanol concentration, a consequence of the fuel's inferior calorific value dominating the combustion characteristics. Conversely, under high load, the BTE surface response is greatly enhanced and widened, with maximum efficiency zone expanded to include mid to high range propanol blends (12 - 18%) and speeds (2600–3300 RPM).

Fig. 15(a) and 15(b) show that the CO emissions decrease with increasing concentration of propanol and engine speed. This reduction can be attributed to the oxygen content of propanol, which favors more complete oxidation. Additionally, increased engine loads have a further effect in reducing CO levels through increased post-combustion oxidation. Propanol fraction is found to be the dominating factor affecting the observed trends of CO emissions.

As shown in Fig. 16(a) and (b), a non-monotonic, dome-shaped trend is exhibited by CO_2 emissions. The curve is seen at its peak at moderate-to-high engine speeds and intermediate propanol concentrations. This is attributed to optimized oxidation conditions. Increased in-cylinder turbulence and intrinsic oxygen content of propanol increase the conversion of CO to CO_2. The reduction of CO_2 at higher propanol levels is

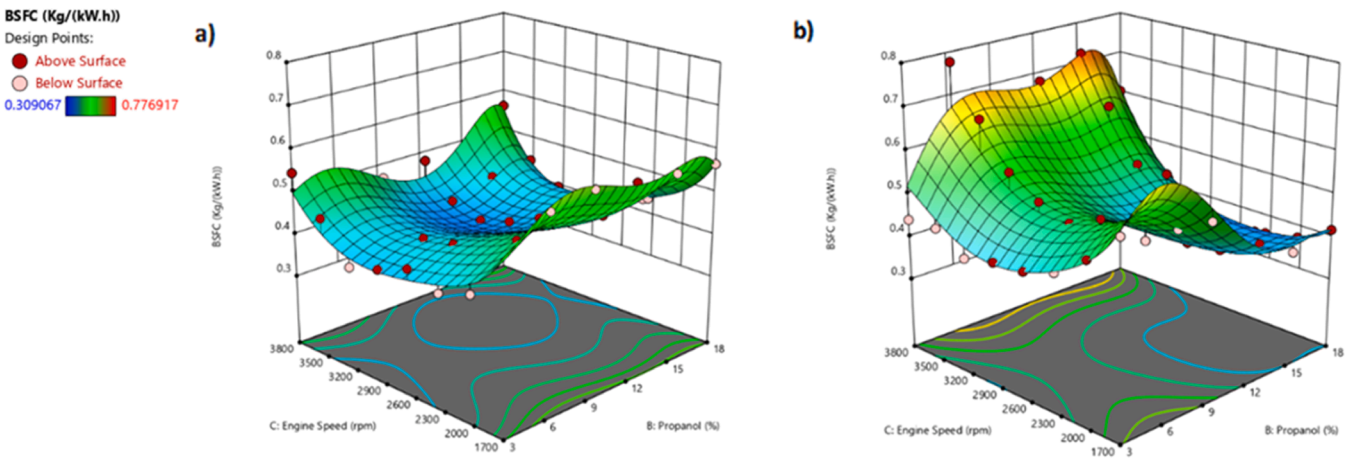

Fig. 13. a) BSFC response curve at low load; b) BSFC response curve at high load.

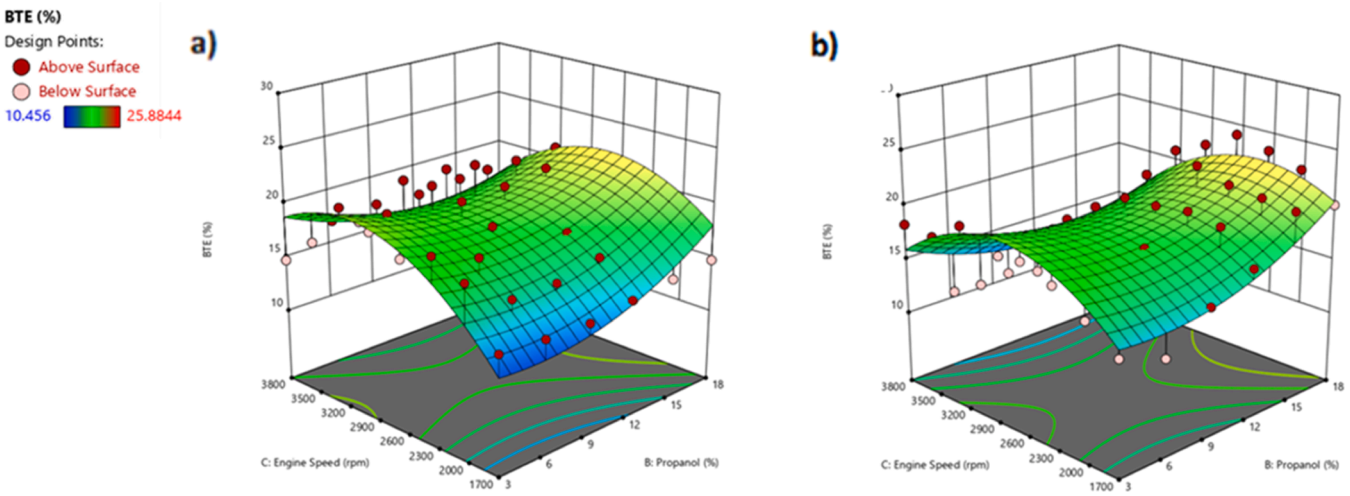

Fig. 14. a) BTE response curve at low load; b) BTE response curve at high load.

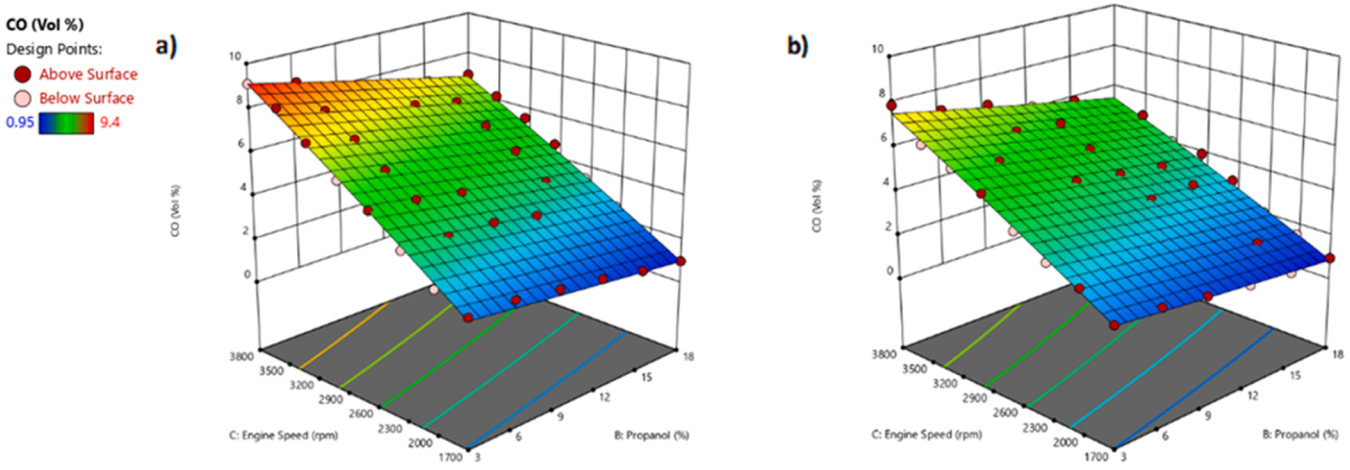

Fig. 15. a) CO response curve for low load; b) CO response curve for high load.

linked with the lower carbon-to-energy ratio of the blend and the intense charge-cooling effect, which in turn slows the combustion process.

CO_2 formation is also limited at lower and upper extremes of speed. At low RPM, poor air–fuel mixing limits complete oxidation. Increasing engine load broadens and elevates the peak, as higher in-cylinder temperatures promote more complete oxidation of carbon species.

Hydrocarbon (HC) emissions, shown in Fig. 17(a) and (b), exhibit an opposite trend. Unburned HC levels decrease consistently with increasing propanol fraction and engine speed. This reduction is attributed to the presence of fuel-bound oxygen and improved mixture

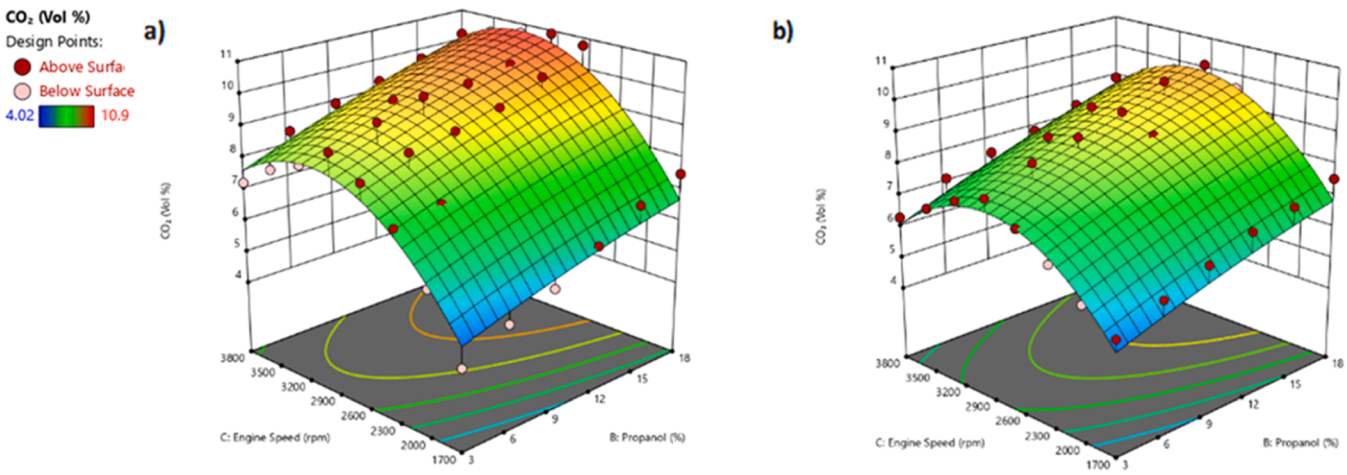

Fig. 16. a) CO_2 response curve at low load; b) CO_2 response curve at high load.

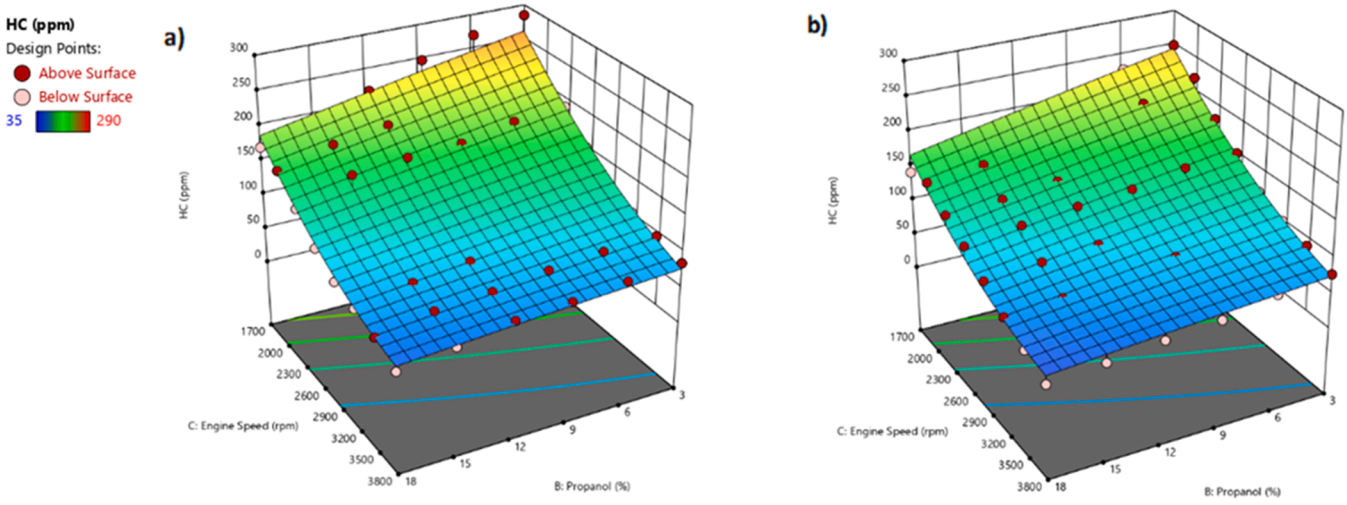

Fig. 17. a) HC response curve at low load; b) HC response curve at high load.

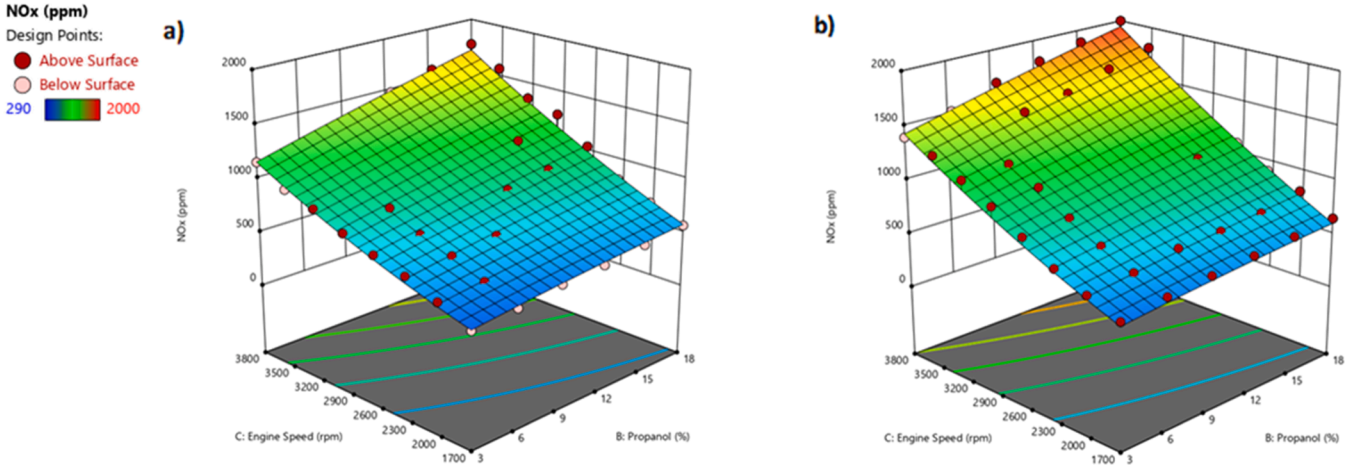

Fig. 18. a) NO_x response curve at low load; b) NO_x response curve at high load.

preparation at higher propanol concentrations, which promote more complete oxidation [35]. Secondly,engines operating at higher speeds increase in-cylinder turbulence - resulting in the persistence of unburned fuel species.

As illustrated in Fig. 18(a) and (b), NOx emissions show an almost-linear behaviour. Emissions appear to be directly proportional to engine speed and inversely proportional to propanol percentage, resulting in a nearly planar RSM curve.

Table 9
P-values for the A-load, B-propanol and C-engine speed.

Response	p-values (A-Load)	p-values (B-Propanol %)	p-values (C-Engine Speed)
Torque	< 0.0001	< 0.0001	< 0.0001
BP	< 0.0001	< 0.0001	< 0.0001
BSFC	0.0041	< 0.0001	0.2599
BTE	0.0008	< 0.0001	0.1589
CO	< 0.0001	< 0.0001	< 0.0001
CO_2	< 0.0001	< 0.0001	< 0.0001
NOx	< 0.0001	< 0.0001	< 0.0001
HC	0.0973	0.0002	< 0.0001

The trends for NO_X are in line with the formation pathways of thermal nitric oxide. This is explained by the extended Zeldovich mechanism. This mechanism is dictated by peak in-cylinder temperatures and availability of oxygen radicals. Eqs. (14–16) show the chemical kinetics fundamentals.

$$N_2 + O \rightleftharpoons NO + N \quad (14)$$

$$N + O_2 \rightleftharpoons NO + O \quad (15)$$

$$N + OH \rightleftharpoons NO + H \quad (16)$$

According to the Zeldovich mechanism, loads increase in-cylinder temperature and produce NO_x formation. On the other hand, additional of propanol in gasoline can suppress thermal NO_x by limiting maximum temperatures through charge-cooling and altering heat-release phasing. Increased turbulence and higher peak temperatures enhance NO_x production, while the reduced residence time at high speeds limits it. The trends observed in Fig. 18 reflect the net effect of these opposing factors.

3.2.2. ANOVA results

ANOVA results in Table 9 confirm that all models are statistically significant ($p < 0.0001$), with engine speed (C) emerging as the dominant factor across most responses. C shows the largest F-values: CO (12,371.96), BP (8364.72), NOx (4105.74), torque (861.72), CO_2 (424.41), and HC (357.51), all with $p < 0.0001$, showing its control over combustion and emissions. Propanol % (B) is the next strongest driver, especially for CO (1686.85), torque (916.59), NOx (750.47), CO_2 (472.40), and BP (588.14) (all $p < 0.0001$), but has a modest effect on HC (15.02, $p = 0.0002$). Load (A) exerts smaller yet significant effects; CO (657.71), NOx (184.14), CO_2 (88.72), torque (47.92) (all $p < 0.0001$), and is not significant for HC (2.82, $p = 0.0973$). Notable exceptions are BSFC and BTE, where engine speed is not significant (BSFC: 1.29, $p = 0.2599$; BTE: 2.02, $p = 0.1589$), while propanol blending remains a primary driver (38.85 and 51.53; both $p < 0.0001$).

Importantly, the statistically significant p-values for the propanol concentration factor (B) across all eight responses ($p < 0.0001$ for seven outputs; $p = 0.0002$ for HC) confirm that the performance improvements attributed to propanol blending (including the increases in torque, brake power, and thermal efficiency, as well as reductions in CO and HC) are statistically significant and are not the result of random experimental variation [41].

ANOVA-based sensitivity (p-values, F-values, and PC%) quantifies factor influence within the assumed polynomial response form and term hierarchy, whereas Gradient Boosting feature importance quantifies the contribution of each input to reducing prediction error in the tree ensemble (capturing nonlinearities and interactions). Therefore, the two measures are not expected to match one-to-one for every response. For example, engine speed can rank highly in ML feature importance because it is strongly predictive across multiple outputs (torque, brake power, and several emissions), even if its main effect is less significant for a specific quadratic ANOVA surrogate (e.g., when its influence is absorbed into interactions or higher-order terms over a constrained domain).

Table 10 presents the percentage contribution (PC%) of each factor and interaction term to the total variation in each response, calculated as the ratio of the term's sum of squares to the corrected total sum of squares. Engine speed (C) dominates BP, CO, and NOx with contributions of 88.60%, 75.70%, and 78.03% respectively, highlighting its overriding influence on power and combustion-related emissions. Propanol blend ratio (B) is the primary driver for two outputs (Torque and CO_2 with contributions of 39.08% and 25.96% respectively), showing the crucial thermochemical impact of alcohol content on combustion efficiency. Load (A) shows moderate contributions across CO and CO_2 (4.02% and 4.88%). For BSFC and BTE, higher-order polynomial terms collectively account for most of the variations (82.73% and 82.17% respectively). This behavior indicates the highly nonlinear nature of these responses across the tested operating range. Residual contributions remain below 6% for all responses, confirming adequate model fit.

The RSM models of BSFC and BTE required higher order polynomial structures to adequately capture their non-linear response behavior observed across the experimental domain. Evaluation of lower-order alternatives, e.g. 2FI models revealed that they did not possess enough explanatory power, warranting the use of the higher order fitting. The aliased nature of fifth order models limits independent estimation of some higher order terms, and consequently, PC% contributions for BSFC and BTE reflect combined term effects rather than isolated factor contributions. This is an inherent constraint of the experimental design resolution and is noted as a limitation of the current RSM analysis.

3.2.3. RSM residual diagnostics

Residual plots for Torque and BSFC were analyzed to further ensure the validity of the RSM models. The normal probability plots of residuals for Torque (Fig. 19a) show points closely following the reference line, proving that the residuals are approximately normally distributed and that the model assumptions are satisfied. The residuals vs. predicted plots for Torque (Fig. 19b) display a random scatter within the ±3.63 boundary with no visible pattern, indicating constant variance and no systematic bias.

For BSFC (Fig. 20a), the residuals follow the reference line in a similar way, well across the central range, though slight deviations are

Table 10
Percentage contribution for each factor.

Source	Torque	BP	BSFC	BTE	CO	CO_2	HC	NOx
A-Load	2.04	0.26	0.68	0.84	4.02	4.88	0.92	3.50
B-Propanol	39.08	6.22	3.02	3.50	10.32	25.96	12.29	14.26
C-Engine Speed	36.75	88.60	0.10	0.14	75.70	23.32	71.10	78.03
AB	0.01	0.01	0.84	0.88	0.04	0.05	0.01	0.01
AC	0.16	—	4.65	5.30	1.03	2.72	0.18	0.95
BC	0.46	1.01	0.37	0.51	0.36	0.31	1.31	1.46
B^2	0.03	0.01	0.83	0.55	—	0.24	—	—
C^2	5.15	0.21	0.86	0.96	—	24.13	—	—
Higher-order/Aliased	11.93	2.60	82.73	82.17	7.89	12.73	—	—
Residual	4.39	1.09	5.92	5.16	0.64	5.66	4.66	2.00

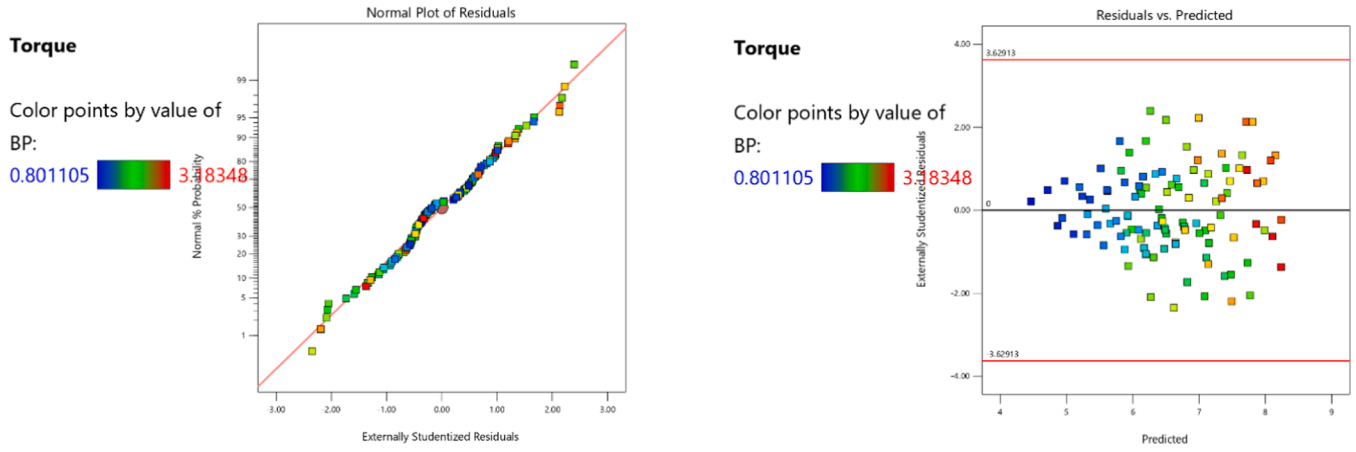

Fig. 19. a) Normal plot of residuals, torque; b) residuals vs. predicted, torque.

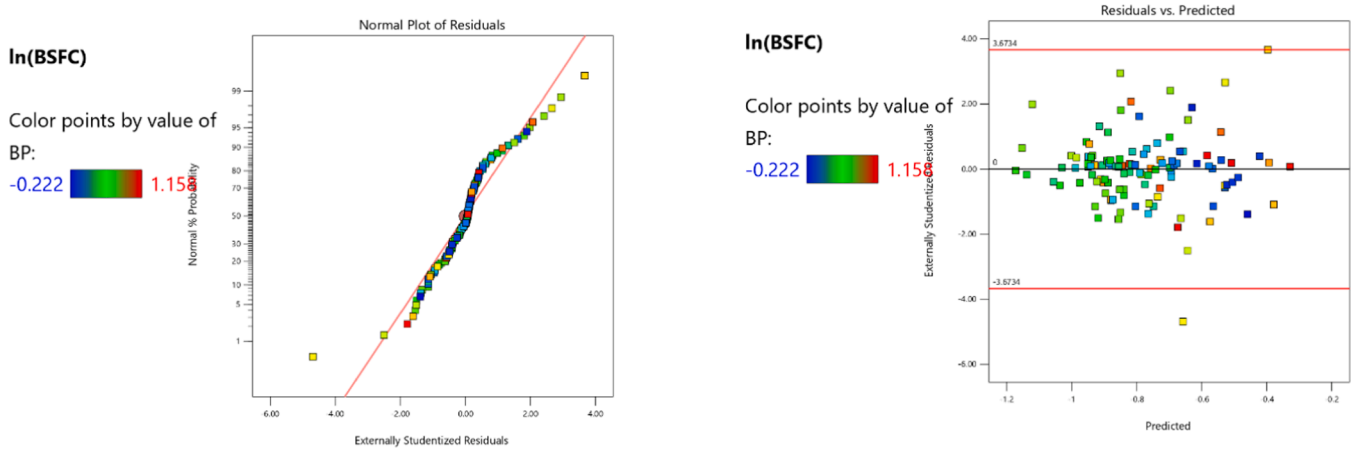

Fig. 20. a) Normal plot of residuals, BSFC; b) residuals vs. predicted, BSFC.

observed at the tails; this is in-line with the variations in BSFC measurement and the moderate R^2 of the RSM model. For BSFC, the residuals vs. predicted plot (Fig. 20b) shows a broadly similar distribution, with one borderline outlier observed below −3.67, consistent with the comparably weaker model fit discussed earlier. Overall, these diagnostics ensure that the RSM models are statistically adequate. Corresponding diagnostic plots for the remaining six responses are provided in the Supplementary Material.

3.3. Multi-objective optimization with NSGA-III

NSGA-III was applied to the three decision variables within the tested region (Load 10–20 psi, Propanol 0–18%, Engine speed 1700–3800 RPM). Eight objectives were optimized simultaneously: minimize BSFC, CO, HC and NO_x; maximize Torque, BP, BTE and CO_2. The evaluations were done using the already trained-and-saved GBM. A Das–Dennis reference-direction set with NSGA-III (population matched to the directions) was evolved for 1200 generations, making a well-spread Pareto front.

Table 11
RSM vs NSGA-III optimum parameters.

Response (unit)	RSM	NSGA-III (knee)	% Difference
Load	20.000	17.936	10.32%
Propanol	18.000	16.628	7.62%
Engine Speed	2930.233	2928.721	0.05%
Torque	8.010	7.926	1.05%
BP	2.457	2.403	2.22%
BSFC	0.370	0.394	6.48%
BTE	22.177	21.972	0.92%
CO	3.600	3.878	7.73%
CO_2	9.893	10.266	3.77%
HC	82.039	87.840	7.07%
NOx	1366.123	1265.443	7.37%

3.3.1. Pareto front and knee solution

The resulting front exhibited the expected trade-offs between performance and emissions. A practical knee point defined as the solution closest to the ideal vector in the normalized objective space was selected as the recommended compromise: Load 17.936 psi, Propanol 16.628%, Speed 2928 RPM. The corresponding responses were Torque 7.926 Nm, BP 2.40 kW, BSFC 0.3939 kg /kW h, BTE 21.97%, CO 3.87 vol %, CO_2 10.26 vol %, HC 87.84 ppm, NO_x 1265.44 ppm. These values lie within the experimental design range and align with the mechanistic expectations from the surface plots.

3.4. Comparison of NSGA-III with RSM desirability optimum

For reference, the RSM desirability search produced a single optimum of similar magnitude. As expected, the RSM point slightly privileged one region of the trade-off space, whereas the NSGA-III knee spread the compromise across objectives which has been shown in Table 11.

Fig. 21 compares the knee-point selection from the Pareto front with the RSM-optimized solution using a normalized radar plot (0–1 scale,

Fig. 21. Radar Plot – RSM vs NSGA-III.

Table 12
Model validation results.

Parameter	Predicted Values	Actual Values	% Error
Torque	8.01	8.10	1.12
BP	2.457	2.484	1.10
BSFC	0.37	0.382	3.33
BTE	22.18	21.92	1.14
CO	3.6	3.77	4.72
CO_2	9.89	10.00	1.08
HC	82.04	85.00	3.61
NO_x	1366.12	1314.00	3.82

where larger values indicate better objective performance). For consistency, objectives to be minimized (BSFC, CO, HC and NO_x) were inverted prior to normalization, while torque, BP, BTE and CO_2 were maximized. The two polygons largely overlap, indicating that the RSM optimum was broadly consistent with the knee-point trade-off.

3.5. Confirmation tests

3.5.1. Experimental model validation

Confirmation tests were conducted to verify the accuracy of the predicted values from the optimization model. Table 12 shows the comparison between predicted and actual experimental results for the 18% propanol blend.

The model shows good accuracy with all parameters having errors below 5%, which is acceptable for engine testing. The highest error of 4.72% for CO emissions is still within the typical measurement uncertainty range for emission testing.

This level of validation accuracy is consistent with confirmation test outcomes reported in analogous RSM studies on alcohol-gasoline SI engines, where prediction errors below 5% were similarly achieved at the RSM-optimized operating point [42].

These values serve as confirmation that the surrogate correctly captures the true engine response surface in the neighborhood of the optimum and that the optimization identified a physically realizable, not surrogate-artefact, operating point.

3.5.2. Performance comparison

In addition to the confirmation tests, the performance of 18% propanol blend against pure gasoline was also done to evaluate the benefits of propanol addition. The results showed significant improvements with the addition of propanol. Engine power increased by 30.76% and torque increased by 29.19%. The oxygen content in propanol led to better combustion, reducing CO by 37.2% and HC by 36.6%. However, fuel consumption increased by 18.59% due to the lower heating value of propanol.

3.6. Model generalization, limitations, computational cost and future work

This study was conducted on a single-cylinder SI engine and within a bounded operating domain (1700–3800 rpm and 10–20 psi) using propanol–gasoline blends up to 18%. Therefore, the optimized solution (including the 18% propanol case) and reported improvements should be interpreted as valid within the tested speed–load region, and should not be extrapolated to part-load/idle, transient conditions, or other engines without additional data and validation. Nevertheless, the proposed GB–RSM–NSGA III workflow is methodologically transferable: given experimental data, the same multi-output surrogate and many-objective optimization strategy can be retrained and applied to other fuels and engine types (including CI engines) to map Pareto-optimal trade-offs. Regarding sustainability, the SDG 7/SDG 13 relevance is framed in terms of the measured ability to identify operating conditions that improve performance indicators and reduce incomplete-combustion emissions (CO/HC) and NO_x within the experimental domain; a full life-cycle CO_2-equivalent assessment and real-driving/transient validation are beyond the present scope and are recommended as future work.

One other important aspect of this research is the computational cost accrued in the analysis. The GB–RSM–NSGA-III pipeline involves three distinct computational stages:

Stage 1 (Gradient Boosting training and hyperparameter optimization): Optuna TPE was run for 100 trials, each comprising a 5-fold cross-

validated MultiOutputRegressor fit. Using parameter n_jobs = −1, this stage completed in approximately 55 min.

Stage 2 (NSGA-III optimization): with a population of 330 individuals evolved for 1200 generations, the total number of surrogate evaluations is 396,000. For this stage, the full 396,000 evaluations completed in under 7 min (each evaluation consists of a single forward pass through the pre-trained GB model and is sub-millisecond)

Stage 3 (RSM): The fitting and desirability optimization in Design-Expert was completed in approximately 10 min. The full end-to-end pipeline therefore requires approximately 72 min of compute time on consumer hardware (without GPU acceleration). After the surrogate is trained, the NSGA-III stage alone takes <10 min, which means that it is a quick way to re-optimize when changes in constraints occur. This is considerably faster than physics-based CFD optimization, where each design-point evaluation typically requires hours on high-performance computing clusters.

It is imperative to note that before deploying the optimized operating conditions in real-engine or production settings, several practical challenges must be acknowledged. The foremost consideration is the compatibility of propanol-gasoline blends with existing fuel-system materials (seals, injectors, and fuel lines). The other considerations include the recalibration requirements for ignition timing and fuel injection systems and the absence of transient-cycle and cold-start validation in the present study. The current surrogate was trained on steady-state averaged data; extension to real-time control applications would require substantially expanded datasets spanning transient operating modes and engine aging effects. These limitations represent natural directions for future research.

4. Conclusion

This study demonstrates the successful application of machine learning-based optimization for propanol-gasoline blends, achieving significant performance and environmental improvements. The results of gradient boosting model exhibited exceptional predictive accuracy and the average error metrics were very low (Training Set: $R^2 = 0.9934$, MAPE = 0.0099; Test Set: $R^2 = 0.9611$, MAPE = 0.0326), which validated the reliability of gradient boosting model for forecasting of engine parameters. Multi-object problem optimization through RSM and Non-dominated Sorting Genetic Algorithm III (NSGA-III) found optimal operating conditions of 20 psi load, 18% propanol blend ratio and engine speed of 2930 RPM. NSGA-III results showed a less-than-5% (4.96%) average deviation with respect to RSM predictions. Under the studied conditions, propanol blending gave significant performance gains, including a 30.76% increase in brake power, an increase of 29.19% of torque, strong reduction of carbon monoxide (37.2%) and hydrocarbon emissions (36.6%). However, brake specific fuel consumption rose by 18.59%, due to propanol's lower calorific value.

This study adds to the UN Sustainable Development Goals, especially SDG 7 (Affordable and Clean Energy) through improved engine efficiency and less fossil fuel dependence and SDG 13 (Climate Action) via reduction in emissions of incomplete combustion products. Confirmation tests further indicate that all response parameters were within error margin of 5% - proving this research's reliability and robustness as an optimization framework for sustainable development of internal combustion engines.

CRediT authorship contribution statement

Muhammad Usman: Supervision, Methodology, Investigation, Conceptualization. **Syed Muhammad Mehdi Ali Rizvi:** Software, Methodology, Investigation. **Muhammad Sher Ali:** Writing – original draft, Validation, Formal analysis. **Fahid Riaz:** Visualization, Supervision, Resources, Investigation. **Muhammad Saeed:** Visualization, Validation, Project administration. **Essam M. Abo-Zahhad:** Writing – review & editing, Visualization, Funding acquisition, Formal analysis, Data curation.

Declaration of competing 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.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.rineng.2026.110203.

Data availability

Data will be made available on request.

References

[1] H. Liu, S. Yu, T. Wang, J. Li, Y. Wang, A systematic review on sustainability assessment of internal combustion engines, J. Clean. Prod. 451 (2024) 141996.
[2] R.D. Reitz, et al., IJER editorial: The future of the Internal Combustion Engine, 21, SAGE Publications Sage UK, London, England, 2020, pp. 3–10.
[3] B. Ashok, A.N. Kumar, A. Jacob, R. Vignesh, Emission formation in IC engines. NOx Emission Control Technologies in Stationary and Automotive Internal Combustion Engines, Elsevier, 2022, pp. 1–38.
[4] S. Singh, M.J. Kulshrestha, N. Rani, K. Kumar, C. Sharma, D. Aswal, An overview of vehicular emission standards, Mapan 38 (1) (2023) 241–263.
[5] A. Joshi, Review of vehicle engine efficiency and emissions, SAE Int. J. Adv. Curr. Pract. Mobil. 1 (2019–01–0314) (2019) 734–761.
[6] S. Winkler, J. Anderson, L. Garza, W. Ruona, R. Vogt, T. Wallington, Vehicle criteria pollutant (PM, NO x, CO, HCs) emissions: how low should we go? NPJ. Clim. Atmos. Sci. 1 (1) (2018) 26.
[7] D.J. Berger, A.D. Jorgensen, A comparison of carbon dioxide emissions from electric vehicles to emissions from internal combustion vehicles, J. Chem. Educ. 92 (7) (2015) 1204–1208.
[8] Z. Yue, H. Liu, Advanced Research On Internal Combustion Engines and Engine Fuels, 16, MDPI, 2023, p. 5940.
[9] H.K. Ng, Advances in internal combustion engines and fuel technologies. BoD–Books On Demand, 2013.
[10] X. Liu, H. Wang, Z. Zheng, J. Liu, R.D. Reitz, M. Yao, Development of a combined reduced primary reference fuel-alcohols (methanol/ethanol/propanols/butanols/n-pentanol) mechanism for engine applications, Energy 114 (2016) 542–558.
[11] Y. Qian, J. Guo, Y. Zhang, W. Tao, X. Lu, Combustion and emission behavior of N-propanol as partially alternative fuel in a direct injection spark ignition engine, Appl. Therm. Eng. 144 (2018) 126–136.
[12] M.U. Kaisan, L.O. Yusuf, I.U. Ibrahim, S. Abubakar, S. Narayan, Effects of propanol and camphor blended with gasoline fuel on the performance and emissions of a spark ignition engine, ACS. Omega 5 (41) (2020) 26454–26462.
[13] M.K. Jamil, et al., Analysis of the impact of propanol-gasoline blends on lubricant oil degradation and spark-ignition engine characteristics, Energies 15 (15) (2022) 5757.
[14] M. Vikneswaran, C. Saravanan, J. Sasikala, P. Ramesh, E.G. Varuvel, Combustion analysis of higher order alcohols blended gasoline in a spark ignition engine using endoscopic visualization technique, Fuel 322 (2022) 124134.
[15] S. Hu, et al., Comparison of physics-based, semi-empirical and neural network-based models for model-based combustion control in a 3.0 L diesel engine, Energies 12 (18) (2019) 3423.
[16] J. Pulpeiro González, K. Ankobea-Ansah, Q. Peng, C.M. Hall, On the integration of physics-based and data-driven models for the prediction of gas exchange processes on a modern diesel engine, Proc. Inst. Mech. Eng. D. J. Automob. Eng. 236 (5) (2022) 857–871.
[17] K. Ankobea-Ansah, C.M. Hall, A hybrid physics-based and stochastic neural network model structure for diesel engine combustion events, Vehicles 4 (1) (2022) 259–296.
[18] M. Zbikowski, A. Teodorczyk, Machine learning for internal combustion engine optimization with hydrogen-blended fuels: a literature review, Energies 18 (6) (2025) 1391.
[19] A. Bhattacharya, P. Majumdar, Artificial intelligence-machine learning algorithms for the simulation of combustion thermal analysis, Heat Transf. Eng. 45 (2) (2024) 176–193.

[20] X. Sun, J. Fu, F. Zhou, B. Luo, J. Liu, Performance prediction and multi-objective optimization for the atkinson cycle engine using eXtreme Gradient Boosting, Therm. Sci. Eng. Prog. 48 (2024) 102402.
[21] O.I. Awad, et al., Response surface methodology (RSM) based multi-objective optimization of fusel oil-gasoline blends at different water content in SI engine, Energy Convers. Manag. 150 (2017) 222–241.
[22] A. Dubey, R.S. Prasad, J.K. Singh, A. Nayyar, Optimization of diesel engine performance and emissions with biodiesel-diesel blends and EGR using response surface methodology (RSM), Clean. Eng. Technol. 8 (2022) 100509.
[23] A.N. Abdalla, et al., Prediction of emissions and performance of a gasoline engine running with fusel oil–gasoline blends using response surface methodology, Fuel 253 (2019) 1–14.
[24] M. Aydın, S. Uslu, M.B. Çelik, Performance and emission prediction of a compression ignition engine fueled with biodiesel-diesel blends: a combined application of ANN and RSM based optimization, Fuel 269 (2020) 117472.
[25] M. Usman, et al., AI-driven optimization of ethanol-powered internal combustion engines in alignment with multiple SDGs: a sustainable energy transition, Energy Convers. Manag. X. 20 (2023) 100438.
[26] M. Aswathanrayan, et al., Prediction of performance and emission features of diesel engine using alumina nanoparticles with neem oil biodiesel based on advanced ML algorithms, Sci. Rep. 15 (1) (2025) 12683.
[27] D.J. Godwin, E.G. Varuvel, M.L.J. Martin, Prediction of combustion, performance, and emission parameters of ethanol powered spark ignition engine using ensemble Least Squares boosting machine learning algorithms, J. Clean. Prod. 421 (2023) 138401.
[28] S. ŞAHİN, T. Eryilmaz, N. Orhan, M. Ertuğrul, Hybrid experimental–ML–RSM framework for optimizing diesel engine performance with waste tire oil blends, Energy (2025) 137491.
[29] A. Ahmad, A.K. Yadav, S. Hasan, Exploring the usability of biohydrogen and biodiesel in internal combustion engines with exhaust gas recirculation: rSM-ANN-based prediction and optimization, Environ. Dev. Sustain. (2025) 1–26.
[30] Y. Singh, A. Sharma, S. Tiwari, A. Singla, Optimization of diesel engine performance and emission parameters employing cassia tora methyl esters-response surface methodology approach, Energy 168 (2019) 909–918.
[31] A. Onawumi, O. Fayomi, S. Okolie, T. Adio, N. Udoye, A. Samuel, Determination of a spark ignition engine's performance parameters using response surface methodology, Energy Procedia. 157 (2019) 1412–1422.
[32] J. Badra, O. Owoyele, P. Pal, S. Som, A machine learning-genetic algorithm approach for rapid optimization of internal combustion engines. Artificial Intelligence and Data Driven Optimization of Internal Combustion Engines, Elsevier, 2022, pp. 125–158.
[33] O. Owoyele, et al., Application of an automated machine learning-genetic algorithm (AutoML-GA) coupled with computational fluid dynamics simulations for rapid engine design optimization, Int. J. Engine Res. 23 (9) (2022) 1586–1601.
[34] S. Sonawane, R. Sekhar, A. Warke, S. Thipse, C. Varma, Forecasting of engine performance for gasoline-ethanol blends using machine learning, J. Eng. Technol. Sci. 55 (3) (2023).
[35] I. Yusri, R. Mamat, W. Azmi, A. Omar, M. Obed, A. Shaiful, Application of response surface methodology in optimization of performance and exhaust emissions of secondary butyl alcohol-gasoline blends in SI engine, Energy Convers. Manag. 133 (2017) 178–195.
[36] M. Kapusuz, H. Ozcan, J.A. Yamin, Research of performance on a spark ignition engine fueled by alcohol–gasoline blends using artificial neural networks, Appl. Therm. Eng. 91 (2015) 525–534.
[37] G. Thodda, V.R. Madhavan, L. Thangavelu, Predictive modelling and optimization of performance and emissions of acetylene fuelled CI engine using ANN and RSM, Energy Sources Recov. Util. Environ. Eff. 45 (2) (2023) 3544–3562.
[38] A.A. Moiz, et al., A machine learning-genetic algorithm (ML-GA) approach for rapid optimization using high-performance computing, SAE Int. J. Commer. Veh. 11 (2018–01–0190) (2018) 291–306.
[39] S. Yang, et al., Multi-objective optimization of exhaust-heat-management strategy for diesel engines based on RSM-NSGA III and VIKOR hybrid algorithm, Energy (2025) 137987.
[40] J.V. Tirkey, D.K. Singh, Thermodynamic performance and emission prediction of CI engine fueled with diesel and Vachellia nilotica (Babul) biomass-based producer gas and optimization using RSM, Pet. Sci. Technol. 40 (9) (2022) 1084–1108.
[41] P. Pugazhendi, G. Dhamodaran, R. Munisamy, M.N. Nachiappan, B.C. Nookaraju, R. Kannan, Experimental and machine learning based investigation of performance and emission characteristics of a CI engine using fusel oil blends, J. Renew. Sustain. Energy. 17 (6) (2025).
[42] M.A. Ijaz Malik, et al., Response surface methodology application on lubricant oil degradation, performance, and emissions in SI engine: a novel optimization of alcoholic fuel blends, Sci. Prog. 106 (1) (2023) 00368504221148342.