

Article

Influence of Renewable *Glycine max* on Engine Performance, Combustion and Emission: A Combined Model Data Validation and Machine Learning Approach

Md Nurun Nabi ¹, Fazlur Rashid ², Nashia Ahmed Nabila ³, Nahina Islam ¹, Roksana Yasmin ¹, Biplob Ray ¹, Inji Kenawy ⁴ and Quddus Tushar ^{1,5,*}

¹ School of Engineering and Technology, Central Queensland University, Melbourne, VIC 3000, Australia; m.nabi@cqu.edu.au (M.N.N.); n.islam@cqu.edu.au (N.I.); r.yasmin@cqu.mail.com (R.Y.); b.ray@cqu.edu.au (B.R.)

² Mechanical and Aerospace Engineering Department, Rajshahi University of Engineering and Technology, Rajshahi 6204, Bangladesh; frrashed10@gmail.com

³ Department of Computer Science, Asian Institute of Technology, Bangkok 12120, Thailand; nashia.ahmed@gmail.com

⁴ School of Computing Engineering and the Built Environment, Edinburgh Napier University (ENU), Edinburgh EH10 5DT, UK; i.kenawy@napier.ac.uk

⁵ School of Engineering, Royal Melbourne Institute of Technology (RMIT), Melbourne, VIC 3053, Australia

* Correspondence: quddus.tushar@rmit.edu.au

Abstract

With increasing energy demand and environmental emission concerns, biodiesels have emerged as a promising alternative to traditional fossil diesel. This study investigates the effects of soybean (*Glycine max*) biodiesel–diesel blends containing 0–80% biodiesel on the performance, combustion, and emission characteristics of a diesel engine. The performance parameters evaluated include brake specific fuel consumption (BSFC), brake mean effective pressure (BMEP), brake power (BP), and brake thermal efficiency (BTE). Combustion characteristics are assessed through in-cylinder pressure, heat release rate (HRR), and cumulative heat release fractions (HRFs), while emission characteristics are analysed in terms of smoke opacity and nitrogen oxides (NO_x) emissions. The primary objective of this study is to evaluate the impact of high-percentage biodiesel and fuel injection timing on overall engine characteristics. A one-dimensional (1D) model was developed using Diesel-RK (Version: 4.30.189) and GT-Suite software (Version: GT-ISE 2026). The models were calibrated and validated through model to model and further validated against experimental data. Using the simulation generated dataset, a machine learning-based surrogate prediction model was developed to rapidly estimate engine performance and emission characteristics directly from engine operating conditions and biodiesel blend ratios, eliminating the need for repeated Diesel-RK simulations. Multiple regression algorithms were evaluated, with the XGBoost tree ensemble model providing the highest prediction accuracy across the investigated performance parameters. Moreover, the Diesel-RK data were validated with experimental data. The results demonstrated model-to-model data and between model to experimental data, confirming the reliability of the developed models. The proposed Diesel-RK simulation and machine learning framework enables rapid and accurate prediction of engine performance and emissions, providing an efficient tool for evaluating biodiesel blends and supporting future engine design and optimisation.



Academic Editor: Manuela Almeida

Received: 15 June 2026

Revised: 15 July 2026

Accepted: 16 July 2026

Published: 17 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: soybean biodiesel; smoke and NOx emissions; performance prediction; simulation; machine learning

1. Introduction

Global energy demand is rising significantly due to advances in technology, population growth, and ongoing development. The relationship between oil prices and the real exchange rate is vital for macroeconomic performance, as fluctuations in oil prices influence exchange rate movements, trade competitiveness, investment conditions, and overall economic stability in both domestic and international markets [1]. Petroleum-based fossil fuels remain one of the primary energy sources, serving a wide range of industrial, transportation, and power generation sectors [2,3]. The combustion of these fuels releases pollutants, including carbon dioxide (CO₂), carbon monoxide (CO), particulate matter (PM), and other pollutants, particularly when used for energy generation and powering automobiles. These greenhouse gases (GHGs) and other pollutants have an adverse impact on human health and the environment. Published studies have raised concerns regarding the abundance of petroleum-based fossil fuels in relation to the current global consumption rates [4]. Another study [5] identified several factors driving the transition toward renewable fuels, including increasing demand for petroleum products, volatility in crude oil prices, growing concerns over national energy security, and uncertainty regarding long-term fuel supply. It is projected that current energy generation practices will be insufficient to sustainably meet future energy demand [6]. For this, finding alternative sources and utilising them as fuel is important [7–13]. Renewable energy sources (biofuel, solar, wind, water, hydro, and so on) are an alternative to mitigating energy demand or to hybridising with fossil fuels, as shown in Figure 1. Renewable fuels would also be suitable for reducing environmental pollution. However, as shown in Figure 1, in 2022, renewable energy accounted for only 7.48%, and hydroelectricity accounted for 6.73% of energy consumption. The other 85.79% was non-renewable energy consumption globally [14]. Hence, focusing on renewable energy sources to meet future energy demand and mitigate GHG emissions is absolutely necessary.

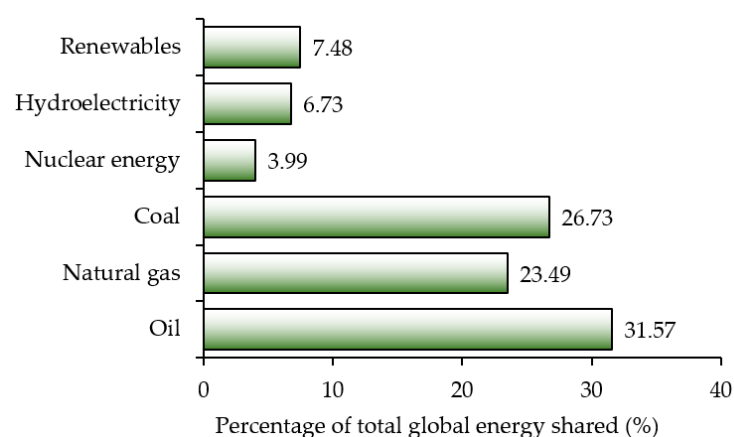


Figure 1. Global energy consumption for the year 2022 by different sources of energy. Adapted from [14].

Transportation sector is one of the largest sectors of using petroleum-based fossil fuels. Despite the growing interest in alternative fuels, conventional fuels such as gasoline, diesel, and natural gas remain widely utilised across road, marine, aviation, and rail transportation sectors. Due to the extensive use of these fuels, GHG emissions, particularly

CO₂ emissions, are primarily the major contributor to anthropogenic climate change. It is noted that atmospheric CO₂ levels were 280 ppm in the pre-industrial era, and have now increased to approximately 430 ppm [15]. Given these impacts on human health and the environment, renewable fuel sources are focused on automobiles and other vehicle fuels. When renewable fuels are considered as fuels for internal combustion engines, it is important to maintain internal combustion engine efficiency, BSFC, brake power, brake torque, BMEP should be comparable to or higher than those realised with conventional fuels, while reducing exhaust emissions. Previous research has focused on the use of different renewable biodiesel fuels in transport vehicles. Biodiesels derived from palm, coconut, and soybean methyl esters were used in diesel engines and found suitable for internal combustion engine fuels [16]. Renewable biodiesel could be one of the ways to abate carbon-related emissions from internal combustion engines. Biodiesels are clean, natural, renewable, efficient, safe to use in diesel engines, non-toxic, and biodegradable [17]. However, due to elevated viscosity, gum formation, acidity, carbon deposits, and free fatty acid content, they are unsuitable for direct use in diesel engines [18]. Biodiesel produced from different vegetable oils, including soybean and sunflower oil, is generally considered cleaner-burning oxygenated fuel. The fuel-bound oxygen in their molecular structure promotes combustion; however, this additional molecular oxygen tends to increase NO_x emissions [19]. Compared to traditional diesel fuel, the higher NO_x emissions with biodiesel can be reduced by employing the exhaust gas recirculation (EGR) technique [20]. The other option to reduce the exhaust emissions is to increase the air-to-liquid ratio in a swirl-type combustor [16]. In an experimental study used an axial flame burner instead of traditional diesel or gasoline engine [16]. The authors conducted the experiments by preheating the main swirl-air, and eventually higher NO_x, CO, and CO₂ emissions were reported [16].

The use of edible oil feedstocks for biodiesel production has raised concerns regarding the food-versus-fuel conflict, motivating the exploration of alternative non-edible feedstocks. Inedible oil feedstocks offer a sustainable and environmentally favourable alternative to edible oils for biodiesel production by reducing competition with food resources and enhancing the potential for non-food-based renewable fuel pathways. Among various non-edible feedstocks, castor oil could be a promising source for biodiesel production. However, castor oil-derived biodiesel generally exhibits slightly higher viscosity than conventional diesel fuel (100D), which may influence fuel atomisation, spray characteristics, and combustion behaviour in diesel engines. In a study [21], the authors conducted experiments in a diesel-fed micro-turbine engine. When castor biodiesel was blended with diesel, the CO was slightly higher, NO_x was lower and there was an absence of SO₂ emissions [21]. Rehman et al. [22] investigated the performance of a gas turbine fuelled with jatropha biodiesel–diesel blends. The results indicated that the incorporation of jatropha biodiesel reduced CO and unburned hydrocarbon (UBHC) emissions. However, it resulted in increased NO_x emissions due to the higher combustion flame temperature. On the other hand, biodiesel from soybean oil showed lower NO_x emissions (about 60% lower) than diesel in micro-gas turbines [23]. Another study found that biodiesel from soybean oil produced higher NO_x emissions because the blended fuel with diesel had larger particles, which take longer to evaporate. This longer period of evaporation led to higher NO_x emissions than diesel [24]. This controversy needs to be clarified for the future use of soybean oil biodiesel. A study [25] reported about 93.3% biodiesel yield from a combination of sunflower and soybean oil. The yield estimation was compared to the experimental biodiesel yield. The authors also estimated BTE, UBHC and NO_x emissions. The response surface methodology (RSM) optimisation results for BTE, UBHC, and NO_x indicated that the deviations of these estimations were 3.34%, 1.35%, and 2.31% respectively. However, their report did not include the details of engine performance parameters

and the systematic analysis of the influence of injection timing on emissions and engine performances. Huang et al. [26] did an experimental investigation with 100% soybean biodiesel and 10% and 20% methanol–biodiesel blends. The authors did experiments with five different engine loads for these fuels. Although the authors reported a decline in CO and UBHC emissions under some operating conditions, their study did not include the effect of engine speed and injection timing on engine performance and emissions. It was also observed that the experimental results were not compared with the traditional petroleum diesel.

On the other hand, the impact of injection pressure on engine emissions, including CO, NO_x, and sulphur dioxide (SO₂), has also been investigated [27]. The authors indicated that an injection pressure of 150 bar provided meaningful performance outcomes and recommended the use of relatively low injection pressures. Engine rotational speed (rpm) is another critical parameter that strongly affects combustion behaviour, engine performance, and exhaust emission characteristics. In a study by Gokalp et al. [28], the authors conducted engine experiments using marine diesel and SME in a diesel-fed tractor engine. The authors examined engine combustion, performance and exhaust emissions under variable engine speed. However, the effects of injection timing on engine operating characteristics were not explored. Hence, this investigation presents the diesel engine performance, combustion and exhaust emissions for soybean methyl ester (SME)/soybean biodiesel blends (up to 80%) under different engine operating conditions, including engine speed and injection timing. The results were compared with a reference diesel fuel (100D). The investigation was conducted with a widely used Diesel-RK simulation tool [29,30]. The study also presents a three-dimensional plot of these engine performance characteristics across different engine speeds and injection timings. Although numerical simulation tools such as Diesel-RK provide detailed insights into engine performance under different operating conditions, repeated simulations across multiple fuel blends, engine speeds, and injection timings can be computationally intensive and time consuming. Recent advances in machine learning (ML) have enabled the development of surrogate prediction models that learn the relationship between engine operating conditions and performance characteristics from simulation or experimental data. Once trained, these models can rapidly estimate engine performance and emissions without repeatedly executing computationally expensive simulations, making them valuable tools for engine optimisation, fuel screening, and decision support. Recent studies have successfully applied machine learning to predict engine energy consumption, combustion characteristics, engine performance, and emissions using both simulation and experimental datasets [31]. Furthermore, recent review studies have highlighted the growing adoption of machine learning, particularly ensemble learning methods, for modelling the complex nonlinear behaviour of internal combustion engines and optimising engine performance and emissions under different fuel conditions [32,33]. However, limited research has integrated high-fidelity Diesel-RK simulations with machine learning to develop surrogate prediction models for SME-blended diesel fuels across a wide range of operating conditions.

This paper investigates the performance, combustion, and emission characteristics of soybean methyl ester (SME)-blended diesel fuels using the Diesel-RK simulation platform under varying engine speeds and fuel injection timings. The simulation results are analysed to evaluate key engine performance parameters, including BSFC, BTE, BMEP, brake BP, as well as NO_x and smoke emissions. Furthermore, the simulation-generated dataset is used to develop machine learning-based surrogate prediction models capable of rapidly estimating these engine performance and emission characteristics directly from engine operating conditions and biodiesel blend ratios. Multiple regression algorithms are evaluated to identify the most accurate prediction model, thereby providing a com-

putationally efficient alternative to repeat Diesel-RK simulations for engine performance assessment and biodiesel blend optimisation. Finally, the Diesel-RK simulation results are validated using GT-Suite and engine experimental results to verify the reliability of the proposed framework.

A recent study [34] has discussed the effectiveness of simulation-based approaches for analysing and optimising energy transformation systems under various operating environments. Systematic modelling and parametric analysis have been widely exercised to explore the thermodynamic characteristics and environmental impacts of developed energy systems. Similarly, this study employs a comprehensive simulation approach, strengthened by experimental validation, to examine the impact of soybean oil methyl ester (soybean biodiesel) on engine performance, combustion characteristics, and exhaust emissions in a compression ignition engine. This paper investigates the performance, combustion, and emission characteristics of SME-blends using the Diesel-RK simulation platform under varying engine operating conditions. The simulation outputs are used to analyse key engine performance parameters, including BSFC, BTE, BMEP, BP, NO_x emissions, and Bosch smoke number. Furthermore, the simulation-generated dataset is used to develop machine learning-based surrogate prediction models capable of rapidly estimating these engine performance and emission characteristics directly from engine operating conditions and biodiesel blend ratios. Multiple regression algorithms are evaluated to identify the most accurate prediction model, thereby providing a computationally efficient alternative to repeat Diesel-RK simulations for engine performance assessment and biodiesel blend optimisation. Finally, the Diesel-RK simulation results are validated using GT-Suite and experimental data to verify the reliability of the proposed framework.

2. Materials and Methods

2.1. Tested Fuels

This study conducted a 1D modelling and simulation investigation of four SME blends (20SME, 40SME, 60SME, and 80SME) along with 100D. Depending on the elemental fuel composition (C, H, O), and blend ratio, fuel properties, including density, kinematic viscosity, oxygen content, lower heating value (LHV), and cetane number were different. Table 1 presents the fuel properties of these four SME blends and 100D.

Table 1. Diesel and SME-blended fuel properties.

Fuels	Abbreviation	Density (kg/m ³)	Kinematic Viscosity (cSt)	LHV (MJ/kg)	Cetane Number	Oxygen (%)	Hydrogen (%)	Carbon (%)
100% Diesel	100D	830	3.61	42.5	48	0	87.00	13.00
20% SME + 80% Diesel	20SME	841	3.82	41.18	48.68	2.59	12.45	84.96
40% SME + 60% Diesel	40SME	852	3.94	39.89	49.37	4.73	12.30	82.97
60% SME + 40% Diesel	60SME	863	3.99	38.73	49.98	6.44	12.35	81.21
80% SME + 20% Diesel	80SME	879	4.10	37.47	50.64	8.61	12.10	79.29

2.2. Tested Engine Specifications

Diesel-RK software [29,30,35] was used to model and simulate engine performance, combustion and emission parameters. For this purpose, a 1D model was developed with this software. A single-cylinder four-stroke direct injection diesel engine with geometric specifications including a bore of 150 mm and a stroke length of 160 mm, a compression

ratio of 20, a fuel injection timing of 13° bTDC and an engine speed of 2000 rpm was chosen for 1D model development. In addition to the engine specifications, key fuel properties, including density, viscosity, cetane number and heating value were also incorporated into the input parameters. The diesel combustion process was modelled by the Diesel-RK multi-zone combustion model. The model was based on the multi-zone approach, in which the in-cylinder process is carried out utilising connected fuel injection, spray evolution, and combustion phases, including heat transfer, gas exchange and emission formation. The NO_x emissions were predicted using the extended Zeldovich mechanism, while the soot emissions were predicted using Razleytsev's phenomenological model. In-cylinder heat transfer was computed by Woschini's heat transfer equation integrated into the software. By varying engine speeds and fuel injection timing, the engine performance parameters BSFC, BMEP, engine efficiency, cylinder pressure, heat release rate, smoke, and NO_x emissions were obtained.

The Diesel-RK simulation results generated under different engine operating conditions were subsequently used to construct a machine learning dataset for predicting engine performance and emission characteristics. The machine learning-based prediction framework is described in Section 2.3.

2.2.1. 1D Modelling with GT-Suite Software

For 1D modelling with GT-Suite, the following continuity, momentum, and energy conservation equations were used [36,37]:

$$\frac{d\dot{m}}{dt} = \frac{dpA + \sum_{boundaries}(\dot{m}u) - 4C_f \frac{\rho u |u|}{2} \frac{dxA}{D} - K_p \left(\frac{1}{2} \rho u |u| \right) A}{dx} \quad (1)$$

$$\frac{D(me)}{dt} = -p \frac{dV}{dt} + \sum_{boundaries}(\dot{m}H) - hA_s(T_{fluid} - T_{Wall}) \quad (2)$$

$$\frac{dm}{dt} = \sum_{boundaries} \dot{m} \quad (3)$$

In Equations (1)–(3), $\dot{m}$ = fluid mass flux, D = diameter (equivalent), h = heat transfer coefficient, dp = the pressure difference through dx , dx = length, A = area, H = total specific enthalpy, T_{wall} = wall temperature, K_p = pressure coefficient, T_{fluid} = fluid temperature, V = volume, ρ = density, C_f = friction factor and u = boundary velocity. The details of the 1D modelling with GT-Suite can be found in [36].

2.2.2. Fundamental Energy and Exergy Parameters

In this section, the equations associated with the estimation of the fuel's fundamental exergy and energy parameters are discussed [38]. The fuel's energy rate was computed based on Equation (4).

$$\dot{E}n_f = \dot{m}_f \times LHV \quad (4)$$

where $\dot{E}n_f$ is the fuel input energy rate (kW), $\dot{m}_f$ is the fuel mass flow rate (kg/s) and LHV is the fuel's lower heating value.

Likewise, the fuel's exergy rate was estimated based on Equation (5).

$$\dot{E}x_f = \dot{m}_f \times \phi \times LHV \quad (5)$$

where $\dot{E}x_f$ is the fuel input exergy rate (kW) and ϕ is the exergy factor, which can be estimated based on Equation (6).

$$\phi = [1.0401 + 0.1728 \frac{h}{c} + 0.0432 \frac{o}{c} + 0.2169 \frac{s}{c} \left(1 - 2.0628 \frac{h}{c} \right)] \quad (6)$$

where h , c , o and s are the fuel's mass fractions of hydrogen, carbon, oxygen and sulphur, respectively.

Exergy efficiency (η_{ex}) was calculated based on Equation (7).

$$\eta_{ex} = \frac{\text{Power output (kW)}}{\dot{E}x_f \text{ (kW)}} \quad (7)$$

2.2.3. Model Validation, Accuracy and Uncertainty Analysis

The Diesel-RK and the GT-Suite software were utilised to carry out 1D modelling for predicting engine performance, combustion and exhaust emission parameters. In addition to the comparison of the model-to-model (Diesel-RK and GT-Suite) variations, the GT-Suite model was also validated with the experimental results. The key parameter for this validation was the in-cylinder pressure.

Among the evaluated machine learning models, the XGBoost Tree Ensemble achieved the highest prediction accuracy, with R^2 values approaching 1.0 for several target variables. While these results indicate excellent agreement with the simulation-generated data, they should be interpreted as surrogate-model performance within the investigated simulation domain rather than evidence of universal predictive capability. Moreover, the main sources of uncertainty of these prediction models with Diesel-RK and GT-Suite were the variations in fuel properties, boundary condition approximations and limitations in representing detailed turbulence–chemistry interactions and chemical kinetics. Overall, the models provide a robust framework for analysing biodiesel-operated engine performance, combustion and emission trends, with acceptable accuracy.

2.3. Machine Learning-Based Prediction Framework

A machine learning-based surrogate prediction framework was developed to rapidly estimate the engine performance and emission parameters obtained from the Diesel-RK simulation results. The purpose of the machine learning model was not to replace the physics-based simulation, but to learn the relationship between engine operating conditions and the corresponding simulated outputs so that the target parameters could be predicted within the investigated operating range without repeatedly running Diesel-RK simulations.

The dataset used for the machine learning analysis was extracted from the Diesel-RK simulation outputs generated for D100 and SME-blended fuels, including SME20, SME40, SME60, and SME80. The dataset contained 8011 data points covering engine speeds from 800 to 2400 rpm. For each fuel case, the machine learning models were trained separately using engine speed as the input variable. The target variables were the corresponding engine performance and emission parameters, including BSFC, BP, BTE, NO_x emissions, and Bosch smoke number.

Before model training, the dataset was randomly partitioned into 70% training data and 30% testing data. In addition, three-fold cross-validation was applied to assess the stability of model performance across different data subsets, reduce dependence on a single train–test split, and minimise the risk of overfitting. The independent test set provided an unbiased evaluation of the predictive performance, while cross-validation assessed the consistency of the learned relationships across multiple data partitions. The random partitioning and cross-validation were performed within the investigated simulation domain to evaluate the interpolation capability of the models. As the dataset was generated from validated Diesel-RK simulations, the developed machine learning models are intended as surrogate predictors within the investigated operating conditions and are not designed for extrapolation beyond the simulated engine speed, biodiesel blend ratios, and injection timing ranges.

Four supervised regression algorithms were implemented in the KNIME Analytics Platform: Linear Regression, Random Forest Regression, XGBoost Linear Ensemble Regression, and XGBoost Tree Ensemble Regression. Linear Regression was used as a baseline model, while Random Forest and XGBoost-based models were selected because of their ability to capture nonlinear relationships between engine speed and the target engine performance or emission parameters. The default hyperparameter settings available in KNIME were used for all machine learning models to maintain a consistent and reproducible modelling workflow (Table 2).

Table 2. Hyperparameter settings of machine learning models.

Model	Hyperparameter	Value
Linear Regression	Algorithm	Ordinary Least Squares (default KNIME settings)
Random Forest	Number of trees	100
	Maximum tree depth	50
XGBoost Linear	Booster	Linear
	Objective	Linear regression
	Number of boosting rounds	100
XGBoost Tree	Booster	Tree
	Objective	Linear regression
	Number of boosting rounds	100
	Learning rate (η)	0.3
	Maximum depth	6
	Gamma	0
	Minimum child weight	1
	Subsample	1.0
	Column sample by tree	1.0
	Column sample by level	1.0
	L1 regularisation (α)	1
	L2 regularisation (λ)	1
	Tree method	Auto
	Grow policy	DepthWise

Model performance was evaluated using the coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and mean signed difference (MSD). The R^2 metric was used to evaluate the goodness-of-fit between the predicted and simulated values, while MAE, MSE, RMSE, and MAPE quantified the prediction errors. The MSD was additionally calculated to assess whether the models exhibited any systematic tendency to overestimate or underestimate the Diesel-RK simulation outputs.

Since the dataset was generated from Diesel-RK simulations, the developed models should be interpreted as surrogate prediction models valid within the investigated engine-speed range and fuel-blend conditions, rather than as models intended for extrapolation beyond the simulated operating domain.

3. Results and Discussion

3.1. Engine Performance, Combustion, and Exhaust Emissions

This section presents the DIESEL-RK software-simulated results for SME blends and diesel fuel under different engine speeds and injection timings. Figure 2 illustrates the BSFC and thermal efficiency (BTE) for different SME-blended fuels (20SME, 40SME, 60SME, and 80SME) and 100D at engine speeds ranging from 800 to 2400 rpm. As shown in Figure 2a, the 20SME blend achieves the lowest BSFC and the highest BTE (approximately 39.6%) at

around 1600 rpm, indicating optimum engine efficiency at this operating condition. At lower and higher engine speeds, the increase in BSFC and corresponding reduction in BTE can be attributed to less efficient fuel conversion, resulting from deviations from the optimum combustion conditions. Similarly, the 40SME, 60SME, and 80SME blends exhibit their lowest BSFC and highest BTE at approximately 1400–1600 rpm, with peak efficiencies ranging from 39.0% to 39.3% (Figure 2b–d), indicating that this speed range provides the optimum operating condition for the SME blends. As shown in Figure 2e, 100D achieves the lowest BSFC (0.213 kg/kWh) and the highest BTE (39.85%) at 1600 rpm. In contrast, BSFC increases with increasing SME blend ratio due to the lower calorific value of soybean methyl ester. This finding is consistent with previously published studies reporting BSFC behaviour at different engine speeds [39].

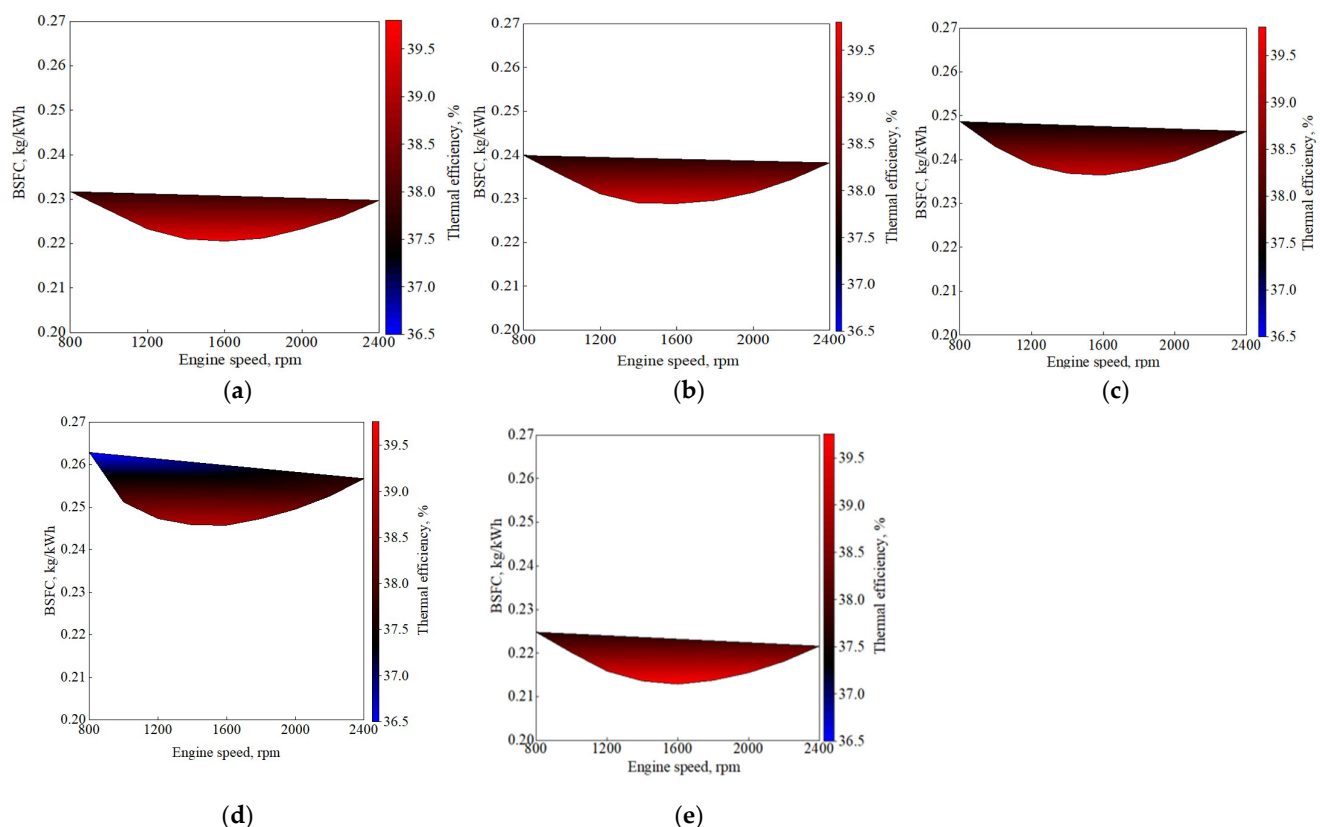


Figure 2. Effect of SME biodiesel on BSFC and thermal efficiency: (a) 20SME, (b) 40SME, (c) 60SME, (d) 80SME, (e) 100D at different engine speeds.

It is generally accepted that fuel consumption is proportional to the volumetric energy density of the fuel, which is determined by its heating value [40]. Fuel consumption increases proportionately with the decrease in LHV and content of biodiesel in blended fuels [41–44]. Overall, the optimum engine speed for minimum BSFC is achieved with improved air–fuel mixing and reduced frictional and pumping losses. Below or above this speed, increased heat transfer or mechanical losses, respectively, contribute to higher BSFC [45].

Compared with 100D, the BSFC increased by 3.6%, 7.4%, 11.2%, and 15.8% for 20SME, 40SME, 60SME, and 80SME, respectively, while the corresponding BTE decreased by 0.43%, 0.8%, 1.47%, and 2.1%. This trend is attributed to the lower energy content of SME, requiring higher fuel consumption to maintain the same output. Similar increases in BSFC for biodiesel blends have been reported in previous studies [46], although negligible BTE variations were observed. For all tested fuels in Figure 3, BTE increased with advancing

injection timing, reached a maximum value, and decreased at further advanced timings, indicating an optimum injection timing for achieving efficient combustion. The maximum BTE and minimum BSFC were achieved within an injection timing range of 9–15° bTDC for the five tested fuels (Figure 3a–e). This range can be considered the optimum injection timing, owing to the inverse relationship between BSFC and thermal efficiency. Compared with 100D, the maximum reductions in BTE for 20SME, 40SME, 60SME, and 80SME were 0.35%, 0.9%, 1.3%, and 2.1%, respectively. The slight decrease in efficiency is attributed to the lower energy content and higher viscosity of SME blends, which adversely affect fuel atomization and combustion characteristics. Published studies [47–49] show lower thermal efficiencies and higher BSFCs with oxygenated fuels and biodiesels compared to traditional diesel fuels. Consistent with the present findings, previous studies have also reported that the reduction in BTE and the increase in BSFC are primarily associated with the lower heating value and higher viscosity of the fuel blends.

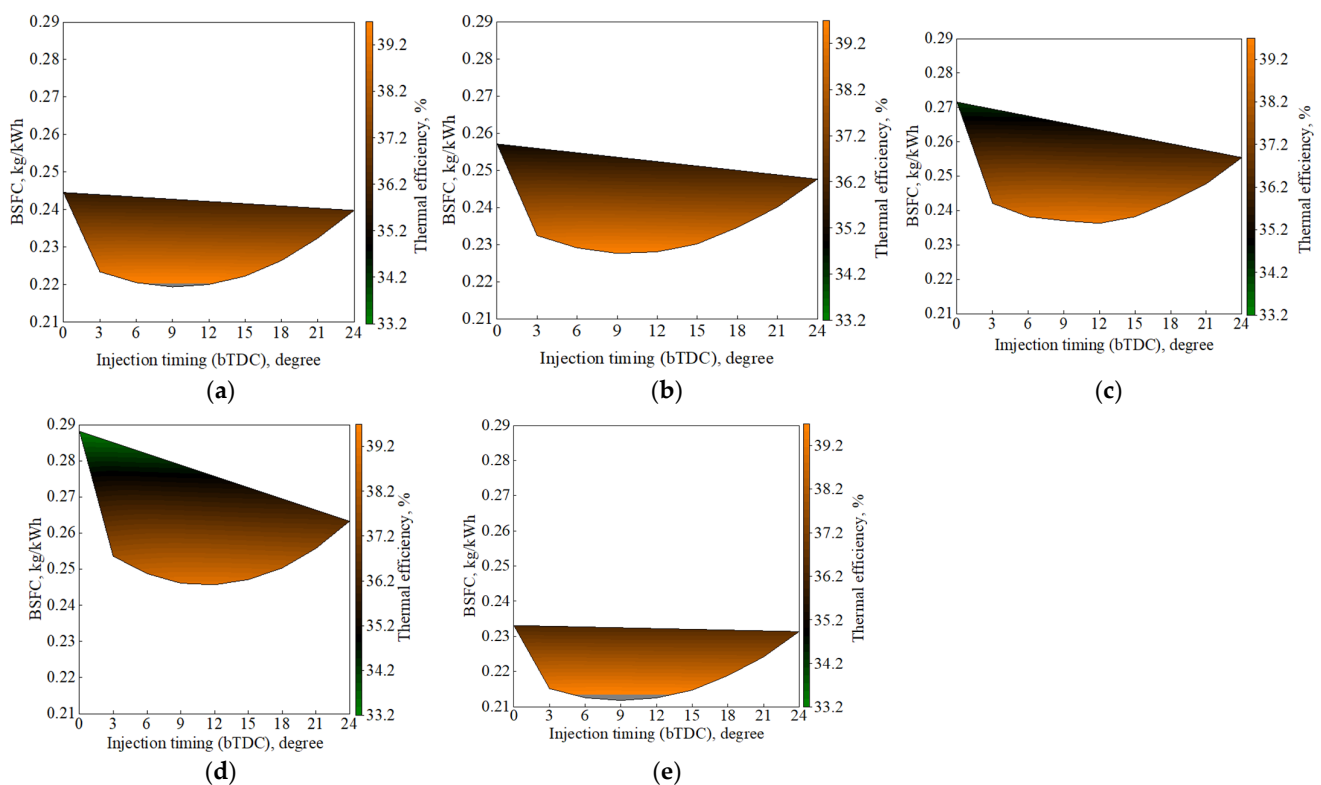


Figure 3. Effect of SME biodiesel on BSFC and thermal efficiency: (a) 20SME, (b) 40SME, (c) 60SME, (d) 80SME, and (e) 100D at different fuel injection timings (bTDC).

The variation in BSFC and BMEP with engine speed in Figure 4a–e shows that low speeds result in higher BSFC and lower BMEP due to increased frictional losses relative to brake power. With increasing speed, BMEP increases while BSFC decreases, reaching optimum performance beyond 1800 rpm. Higher SME proportions reduce BMEP and increase BSFC due to the higher density and lower heating value of SME blends compared with 100D. As demonstrated in Figure 4a–e, the incorporation of SME into diesel fuel resulted in a progressive reduction in BMEP, with decreases of 0.9%, 2.0%, 3.5%, and 4.4% observed for 20SME, 40SME, 60SME, and 80SME, respectively, compared with 100D. A previous study has demonstrated that the use of oxygenated fuels and their blends in compression ignition engines generally results in reduced BMEP and increased BSFC, primarily due to their lower energy density compared with conventional diesel fuel [49].

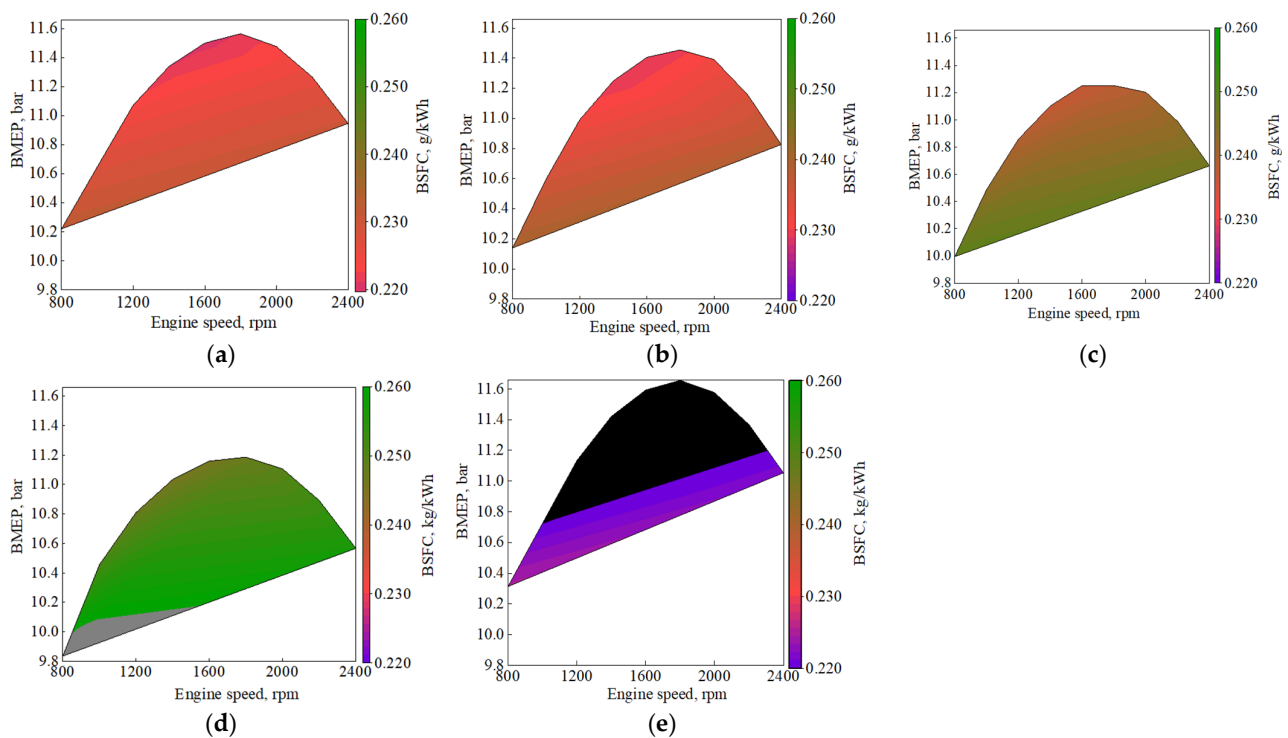


Figure 4. Effect of SME biodiesel on BMEP and BSFC: (a) 20SME, (b) 40SME, (c) 60SME, (d) 80SME, and (e) 100D at different engine speeds.

The in-cylinder pressure variation of 100D and SME blends (20SME–80SME) in Figure 5a was evaluated at a compression ratio of 20 and injection timing of 13° bTDC. The cylinder pressure increased with crank angle advancement, reaching a peak at approximately 6° after the start of combustion, followed by a gradual decline during the expansion stroke. All SME blends exhibited lower peak cylinder pressure than 100D, with 80SME showing the lowest and 20SME the highest pressure among the blends. This reduction in peak pressure is attributed to the lower LHV of SME blends with increasing biodiesel content. Increasing the injection spray angle from 50° to 70° increased the peak cylinder pressure for both 100D and 40SME (Figure 5b). A larger spray angle enhances fuel–air mixing by improving spray distribution within the combustion chamber, resulting in faster combustion and higher cylinder pressure and HRR.

Consistent with the cylinder pressure behaviour, the HRR increased with crank angle and reached a peak at approximately 12° before declining during the expansion stroke (Figure 6a,b). The HRR profiles of 100D and 40SME showed similar trends, with 100D exhibiting a higher peak HRR than 40SME for all injection spray angles. The higher HRR with an appropriate spray angle is attributed to improved fuel–air mixing, which promotes faster combustion and increases peak cylinder pressure [50]. All SME blends exhibited similar ignition delays in Figure 6c, which were shorter than that of 100D due to their higher cetane number (Table 1). The shorter ignition delay reduces fuel accumulation before combustion, resulting in lower HRR peaks. Additionally, the lower energy content of SME blends contributes to the reduced heat release compared with diesel fuel.

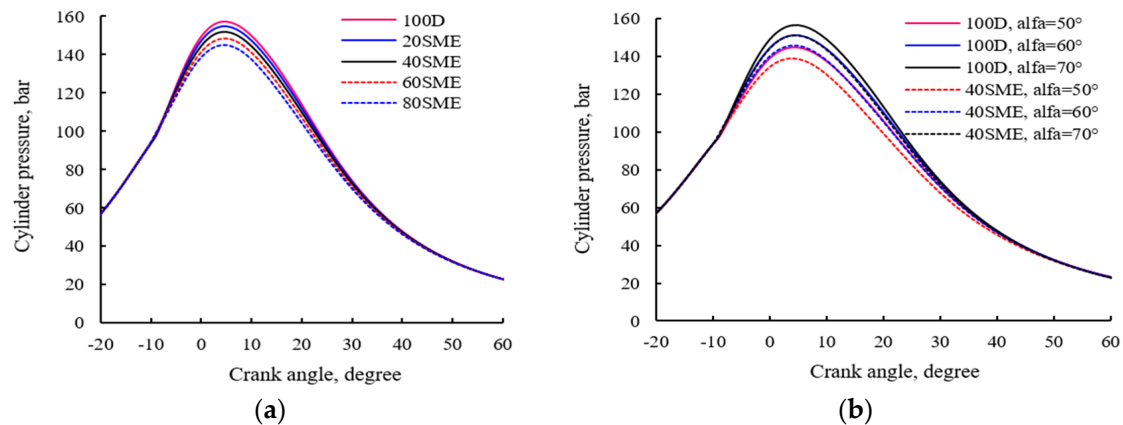


Figure 5. Cylinder pressure at different crank angles for (a) 20SME, 40SME, 60SME, 80SME, and 100D. (b) Influence of spray angle on cylinder pressure for 40SME and 100D.

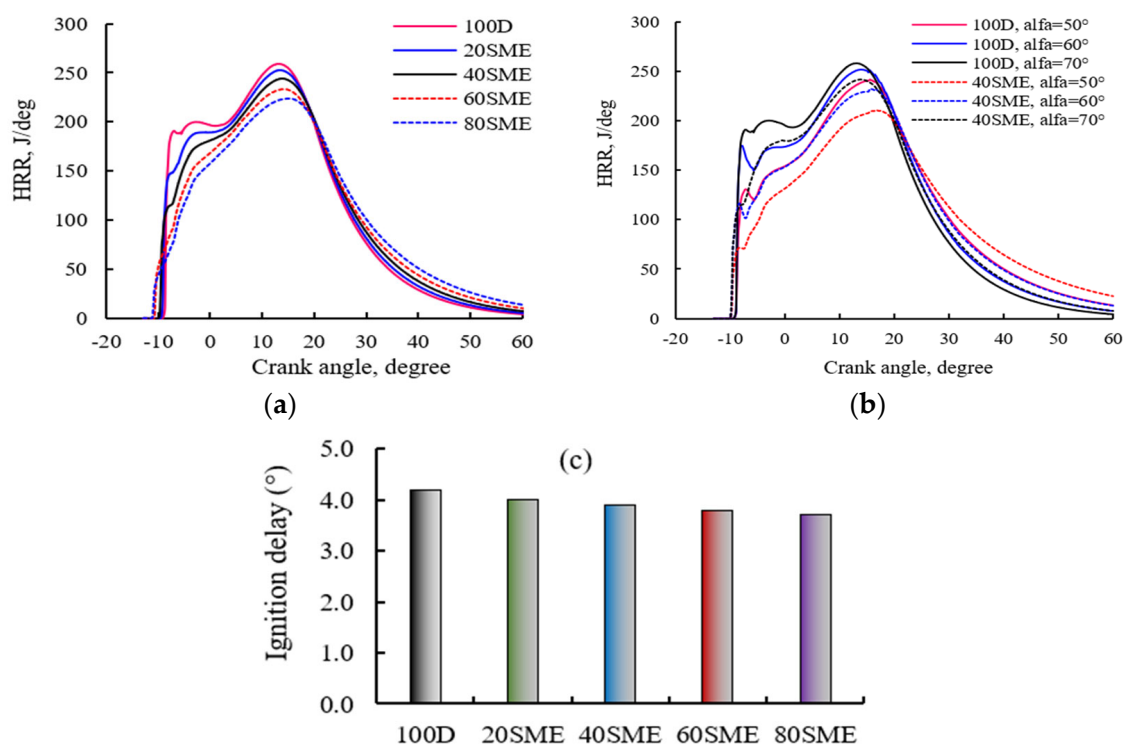


Figure 6. Heat release rate (HRR) at different crank angles of (a) 20SME, 40SME, 60SME, 80SME, and 100D. (b) HRR of 20SME, 40SME, 60SME, 80SME, and 100D at injection spray angles of 50°, 60°, and 70°. (c) Ignition delay for 20SME, 40SME, 60SME, 80SME, and 100D.

Next, the smoke and NO_x emission characteristics of different SME blends and 100D are quantified and presented in Figure 7a–e. In this case, the compression ratio was set at 20 and engine rotational speed was kept constant at 1600 rpm. It is evident from Figure 7a–e that the increase in NO_x emissions, associated with the higher oxygen content of the SME blends, corresponds to a reduction in smoke emissions. The lowest BSFC was observed within the operating region, where NO_x emissions ranged from 6.0 to 12 g/kWh, indicating a relationship between combustion efficiency and NO_x formation characteristics. Also, as displayed in Figure 7a–e, NO_x is higher for higher oxygenated fuels, while the smoke emissions are lower for the same oxygenated fuels (SME blends) compared to 100D. In this case 80SME shows the highest NO_x emissions with the lowest smoke emissions. The findings of the present investigation are in good agreement with the

published literature [51], which report that fuel-borne oxygen plays a key role in reducing smoke emissions.

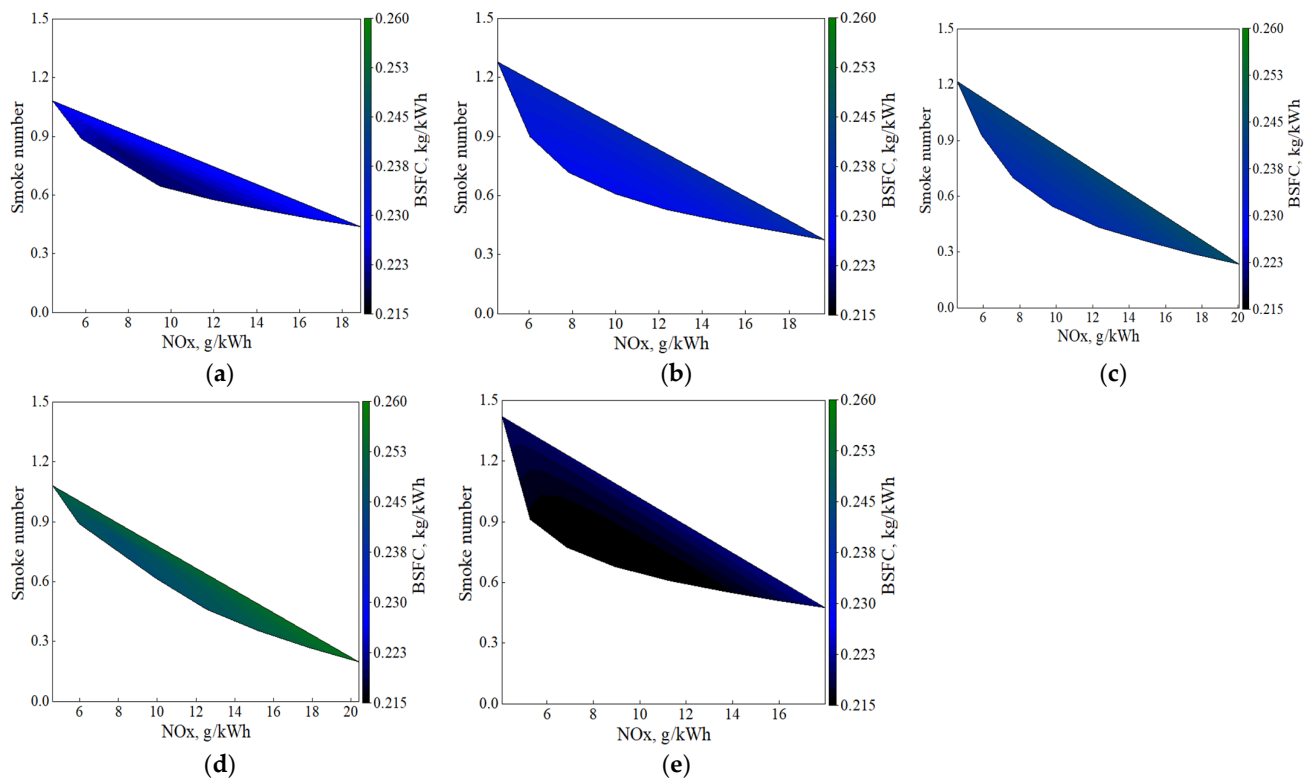


Figure 7. Effect of SME biodiesel on smoke emission and BSFC: (a) 20SME (b) 40SME (c) 60SME (d) 80SME (e) 100D at different levels of nitrogen oxide (NOx).

Figure 8 presents NOx and smoke/soot generation for 100D and 40SME blended fuels under different injection spray angles at a compression ratio of 20 and an injection timing of 13° bTDC. It is observed from Figure 8a that with the increase in injection spray angle from 50° to 70°, the NOx emissions increase for both 100D and 40SME fuels. The increased NOx formation is attributed to higher combustion temperatures, which are strongly influenced by the HRR characteristics shown in Figure 6a,b. The higher HRR at the corresponding injection spray angles promotes elevated in-cylinder temperatures, leading to increased NOx emissions. This higher NOx formation at elevated temperatures has been widely reported in previously published studies [48,52].

The smoke emission is initially low from approximately 5° bTDC to 7° aTDC, increases with crank angle, reaches a maximum at around 12° aTDC, and subsequently decreases, as shown in Figure 8b. Smoke emissions were lower at smaller injection spray angles and increased with increasing spray angles for both 100D and 40SME blended fuels. The higher smoke emissions with increased injection spray angle could be associated with poor air-to-fuel mixing and incomplete combustion zones. All SME blends and 100D display the same trends; however, a closer look at Figure 8b reveals SME blends produce less smoke emissions compared to 100D. The molecular oxygen in SME blends enhanced complete combustion, resulting in reduced smoke emissions. Relative to 100D, 20SME reduced smoke emissions by 7.8%, followed by 21.2%, 50.7%, and 58.5% reductions for 40SME, 60SME, and 80SME, respectively. A previous study [46] reported a 45% smoke reduction with 70% biodiesel and 60% reduction with 100% biodiesel. On the other hand, 20SME increased NOx emissions by 4.8%, while 40SME, 60SME, and 80SME increased emissions by 9.2%, 11.4%, and 13.2%, respectively. Buyukkaya [46] did an engine experiment with different

biodiesel blends and reported a 12% increase in NO_x emissions with 100% biodiesel and 9% increase with 70% biodiesel.

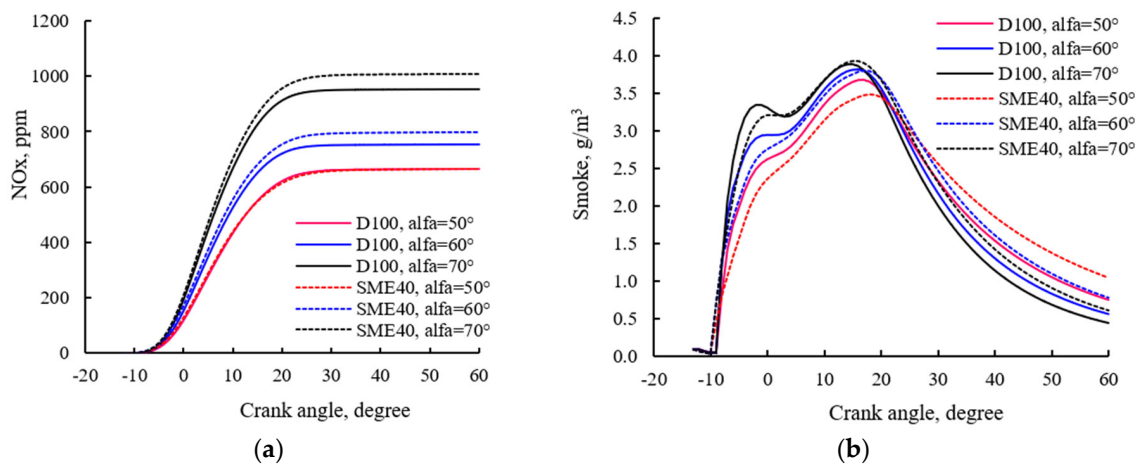


Figure 8. (a) Generation of NO_x at different crank angles for 100D and 40SME biodiesel fuel at injection spray angles of 50°, 60°, and 70°. (b) Soot generation at different crank angles for 100D and 40SME biodiesel fuel at injection spray angles of 50°, 60°, and 70°.

3.2. Performance Analysis of Machine Learning-Based Prediction Model

The machine learning framework was developed to evaluate the feasibility of rapidly predicting engine performance and emission characteristics directly from engine speed using the Diesel-RK simulation dataset. Four supervised regression algorithms, namely Linear Regression, Random Forest Regression, XGBoost Linear Ensemble Regression, and XGBoost Tree Ensemble Regression, were evaluated using the same training and testing datasets to identify the most suitable surrogate prediction model. The models were trained using 70% of the dataset and tested using the remaining 30%, while three-fold cross-validation was employed to assess the robustness and generalisation capability of each algorithm. KNIME Analytics Platform was used for this analysis [53].

Figure 9 compares the predicted BSFC obtained from the four regression models with the corresponding Diesel-RK simulation results for 100D and SME-blended fuels (SME20, SME40, SME60, and SME80). Overall, all four models successfully captured the nonlinear relationship between engine speed and BSFC. The predicted curves closely followed the simulation results, with BSFC decreasing as engine speed increased to approximately 1600 rpm before increasing again at higher engine speeds. This trend is consistent with the Diesel-RK simulation results presented in Section 3.1 and confirms that the machine learning models successfully learned the underlying relationship between engine speed and engine performance. Among the evaluated models, the XGBoost Tree Ensemble Regression model consistently produced predictions that most closely matched the Diesel-RK simulation outputs across all biodiesel blends. The improved prediction accuracy can be attributed to the gradient boosting framework, which effectively models complex nonlinear relationships through sequential optimisation of decision trees. In comparison, the Linear Regression model was limited by its assumption of linearity, while the Random Forest and XGBoost Linear Ensemble models demonstrated good predictive capability but slightly larger deviations from the simulation results in regions where BSFC changed rapidly with engine speed.

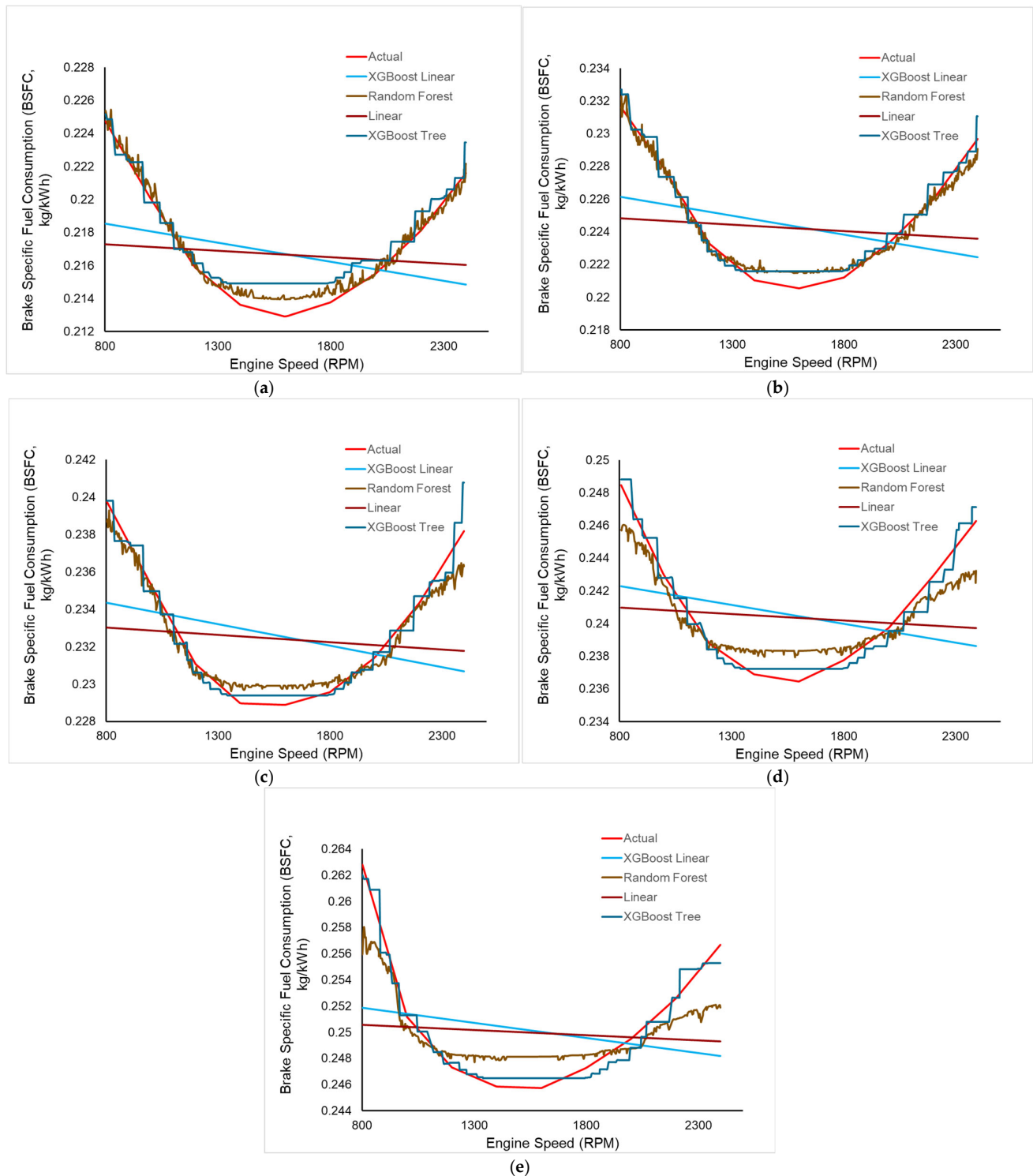


Figure 9. BSFC prediction using machine learning models for (a) 100D, (b) 20SME, (c) 40SME, (d) 60SME and (e) 80SME.

The close agreement between the XGBoost Tree Ensemble predictions and the Diesel-RK simulation demonstrates that the proposed surrogate model can accurately reproduce the simulated engine behaviour within the investigated operating range. Once trained, the model provides rapid predictions without repeatedly executing computationally intensive Diesel-RK simulations, making it suitable for preliminary engine performance assessment, biodiesel blend evaluation, and future optimisation studies.

To quantitatively compare the regression models, several statistical performance metrics were calculated, including the coefficient of determination (R^2), adjusted R^2 , MAE, MSE, RMSE, mean signed difference, and MAPE. Higher values of R^2 and adjusted R^2 , together with lower prediction errors, indicate better agreement between the machine learning predictions and the Diesel-RK simulation outputs. These metrics provide a comprehensive evaluation of each model's predictive capability and generalisation performance.

Table 3 quantitatively compares the performance of the four regression models for BSFC prediction using the statistical metrics described in Section 2.3. The XGBoost Tree Ensemble Regression model consistently achieved the highest R^2 values, together with the lowest MAE, MSE, RMSE, MAPE, and absolute MSD across all fuel blends. These results demonstrate that the model accurately captured the nonlinear relationship between engine speed and BSFC while exhibiting minimal systematic prediction bias. Although the Random Forest model also achieved high predictive accuracy, its error metrics were consistently higher than those of the XGBoost Tree Ensemble model. In contrast, the Linear Regression and XGBoost Linear Ensemble models showed larger prediction errors because they were less capable of representing the nonlinear characteristics of the Diesel-RK simulation data. Overall, the statistical evaluation confirms that the XGBoost Tree Ensemble model is the most suitable surrogate model for BSFC prediction within the investigated operating range.

Table 3. Performance matrices for BSFC of different machine learning models.

Parameters	Linear Regression	Random Forest Regression	XGBoost Linear Ensemble Regression	XGBoost Tree Ensemble Regression
R^2	0.92	0.99	0.918	0.996
Mean absolute error	0.003	0.001	0.003	0.001
Mean squared error	9×10^{-6}	1×10^{-6}	9×10^{-6}	1×10^{-6}
Root mean squared error	0.003	0.001	0.003	0.001
Mean signed difference	0	0	0	0
Mean absolute percentage error	0.013	0.004	0.013	0.003

The machine learning framework was subsequently applied to predict BP. Figure 10 compares the predicted BP obtained using the four regression models with the corresponding Diesel-RK simulation results for 100D and SME20, SME40, SME60, and SME80 fuels. All models successfully reproduced the increasing trend of brake power with engine speed; however, noticeable differences were observed in prediction accuracy. The XGBoost Tree Ensemble model produced predictions that almost completely overlapped with the Diesel-RK simulation curves across all fuel blends, indicating excellent agreement with the simulated data. The Random Forest model also provided accurate predictions, whereas the Linear Regression and XGBoost Linear Ensemble models exhibited slightly larger deviations, particularly at higher engine speeds where the engine behaviour became more nonlinear. The statistical performance metrics presented in Table 4 further confirm the superiority of the XGBoost Tree Ensemble model for BP prediction. This model achieved the highest R^2 , together with the lowest prediction errors (MAE, MSE, RMSE, and MAPE), while maintaining an MSD close to zero, indicating negligible systematic prediction bias. The excellent agreement between the predicted and simulated brake power demonstrates that the proposed surrogate model can accurately estimate engine output across different biodiesel blends without repeatedly executing Diesel-RK simulations.

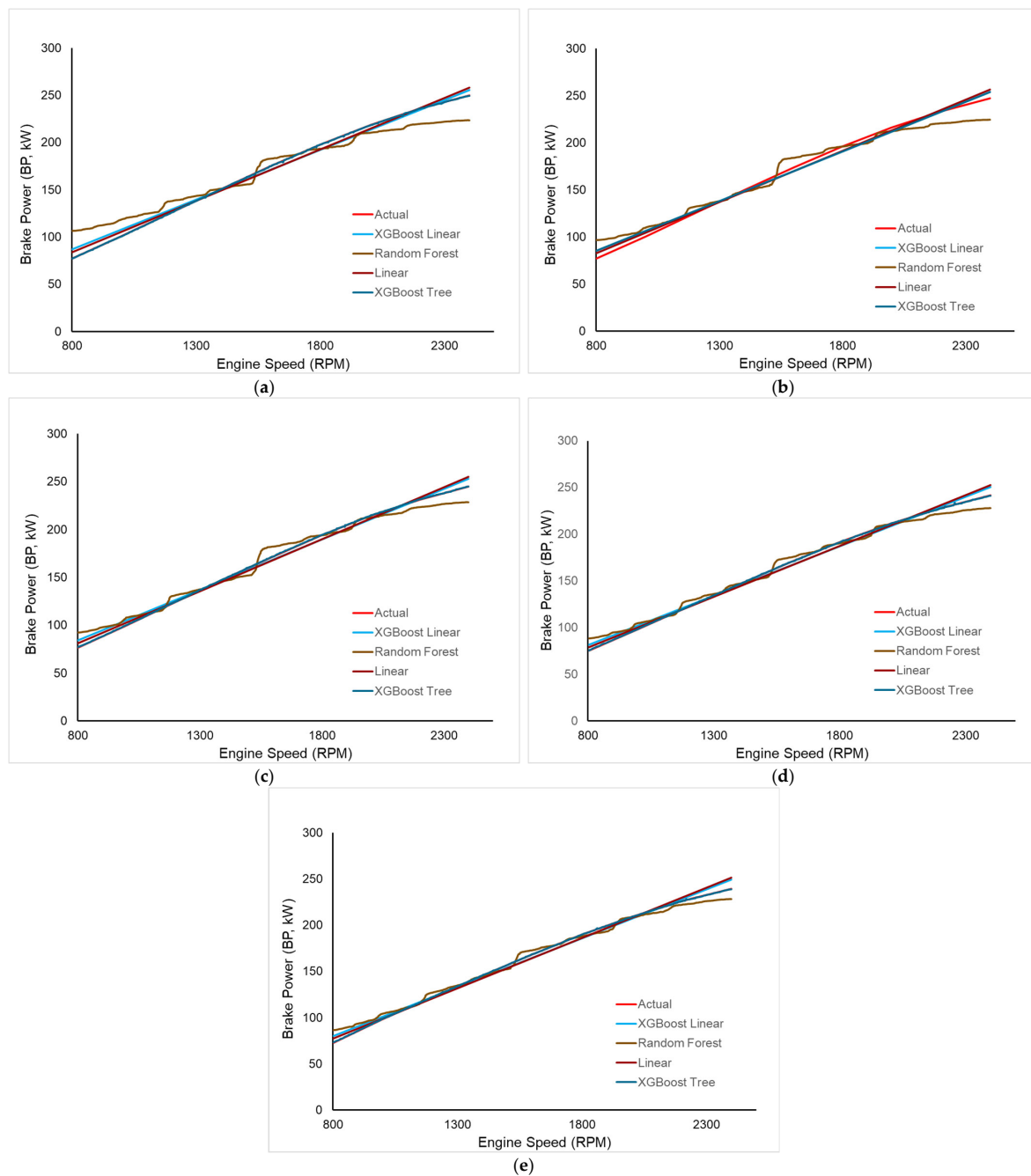


Figure 10. Brake power prediction using machine learning models for (a) 100D, (b) 20SME, (c) 40SME, (d) 60SME and (e) 80SME.

Next, the proposed regression models were evaluated for predicting brake thermal efficiency (BTE). Figure 11 compares the predicted BTE obtained from the four machine learning algorithms with the corresponding Diesel-RK simulation results. Similar to the BSFC analysis, the Random Forest and XGBoost Tree Ensemble models accurately captured the nonlinear variation in BTE with engine speed, including the peak efficiency observed at approximately 1600 rpm. In contrast, the Linear Regression and XGBoost Linear Ensemble models produced larger deviations because the relationship between engine speed and BTE is inherently nonlinear. Among all evaluated algorithms, the XGBoost Tree Ensemble model most closely reproduced the Diesel-RK simulation results across all biodiesel blends.

Table 4. Performance matrices for BP of different machine learning models.

Parameters	Linear Regression	Random Forest Regression	XGBoost Linear Ensemble Regression	XGBoost Tree Ensemble Regression
R^2	0.994	0.97	0.994	1
Mean absolute error	3.149	6.68	3.442	0.171
Mean squared error	14.274	76.655	16.033	0.057
Root mean squared error	3.778	8.755	4.004	0.24
Mean signed difference	0.079	−0.19	0.234	0.004
Mean absolute percentage error	0.02	0.046	0.024	0.001

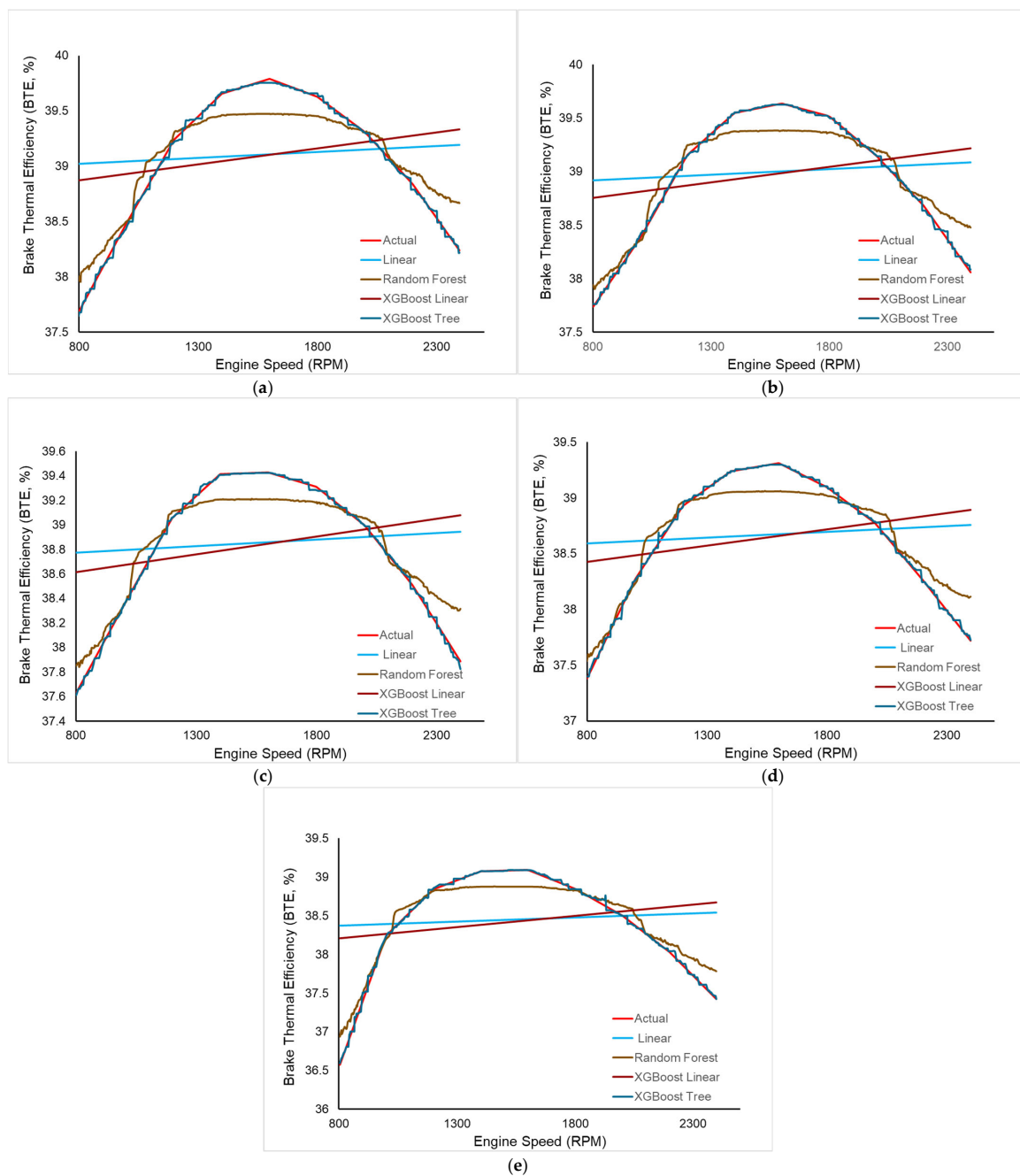
**Figure 11.** BTE prediction using machine learning models for (a) 100D, (b) 20SME, (c) 40SME, (d) 60SME and (e) 80SME.

Table 5 demonstrates that the XGBoost Tree Ensemble model consistently outperformed the remaining regression models for BTE prediction. The combination of high R^2 and low prediction errors confirms that the proposed surrogate model accurately predicts engine thermal efficiency while maintaining minimal systematic bias. These results further demonstrate the capability of gradient-boosted decision trees to model the complex nonlinear relationships associated with biodiesel engine performance.

Table 5. Performance matrices for BTE of different machine learning models.

Parameters	Linear Regression	Random Forest Regression	XGBoost Linear Ensemble Regression	XGBoost Tree Ensemble Regression
R^2	0.155	0.932	0.137	0.999
Mean absolute error	0.477	0.132	0.488	0.016
Mean squared error	0.312	0.025	0.319	0.001
Root mean squared error	0.559	0.159	0.565	0.023
Mean signed difference	0.009	0.005	−0.004	−0.001
Mean absolute percentage error	0.012	0.003	0.013	0

Additionally, the machine learning framework was applied to predict NOx emissions generated by the Diesel-RK simulations. Figure 12 compares the predicted NOx emissions obtained using the four regression algorithms with the corresponding simulation results for 100D and SME-blended fuels. The Random Forest and XGBoost Tree Ensemble models accurately reproduced the nonlinear variation in NOx emissions with engine speed, whereas the Linear Regression and XGBoost Linear Ensemble models showed larger deviations. The XGBoost Tree Ensemble model provided the closest agreement with the Diesel-RK simulations throughout the engine-speed range, demonstrating its ability to model complex emission behaviour.

The statistical evaluation presented in Table 6 confirms that the XGBoost Tree Ensemble model achieved the best overall performance for NOx prediction. Its high R^2 values, together with the lowest MAE, MSE, RMSE, MAPE, and near-zero MSD, indicate both excellent prediction accuracy and minimal systematic bias.

Table 6. Performance matrices for NOx emissions of different machine learning models.

Parameters	Linear Regression	Random Forest Regression	XGBoost Linear Ensemble Regression	XGBoost Tree Ensemble Regression
R^2	0.321	0.964	0.317	0.999
Mean absolute error	0.4	0.098	0.407	0.016
Mean squared error	0.426	0.023	0.429	0.001
Root mean squared error	0.653	0.15	0.655	0.024
Mean signed difference	−0.036	−0.006	−0.039	−0.001
Mean absolute percentage error	0.041	0.01	0.042	0.002

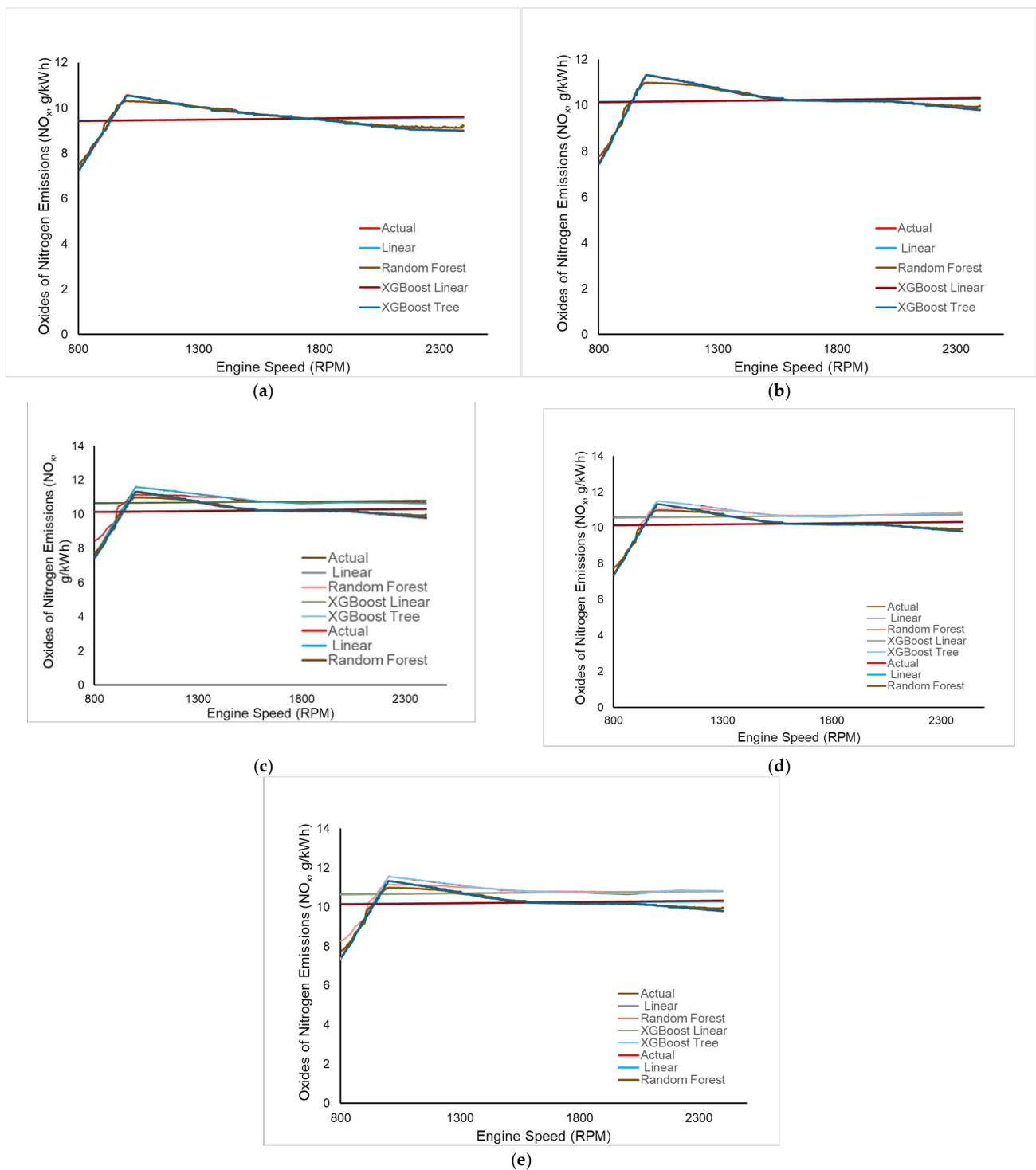


Figure 12. NO_x prediction using machine learning models for (a) 100D, (b) 20SME, (c) 40SME, (d) 60SME and (e) 80SME.

Finally, Figure 13 compares the Bosch smoke number predicted by the four machine learning regression models with the corresponding Diesel-RK simulation results for 100D and SME-blended fuels (20SME, 40SME, 60SME, and 80SME). Overall, all four regression models successfully captured the variation in smoke number with engine speed across the investigated biodiesel blends. The predicted curves closely followed the Diesel-RK simulation results, demonstrating that the machine learning models effectively learned the relationship between engine speed and smoke emissions. The XGBoost Tree Ensemble model exhibited the closest agreement with the Diesel-RK simulation curves throughout

the investigated engine-speed range, whereas the Linear Regression and XGBoost Linear Ensemble models showed comparatively larger deviations, particularly in regions where smoke emissions changed more rapidly. The Random Forest model also achieved high prediction accuracy but was slightly less accurate than the XGBoost Tree Ensemble model.

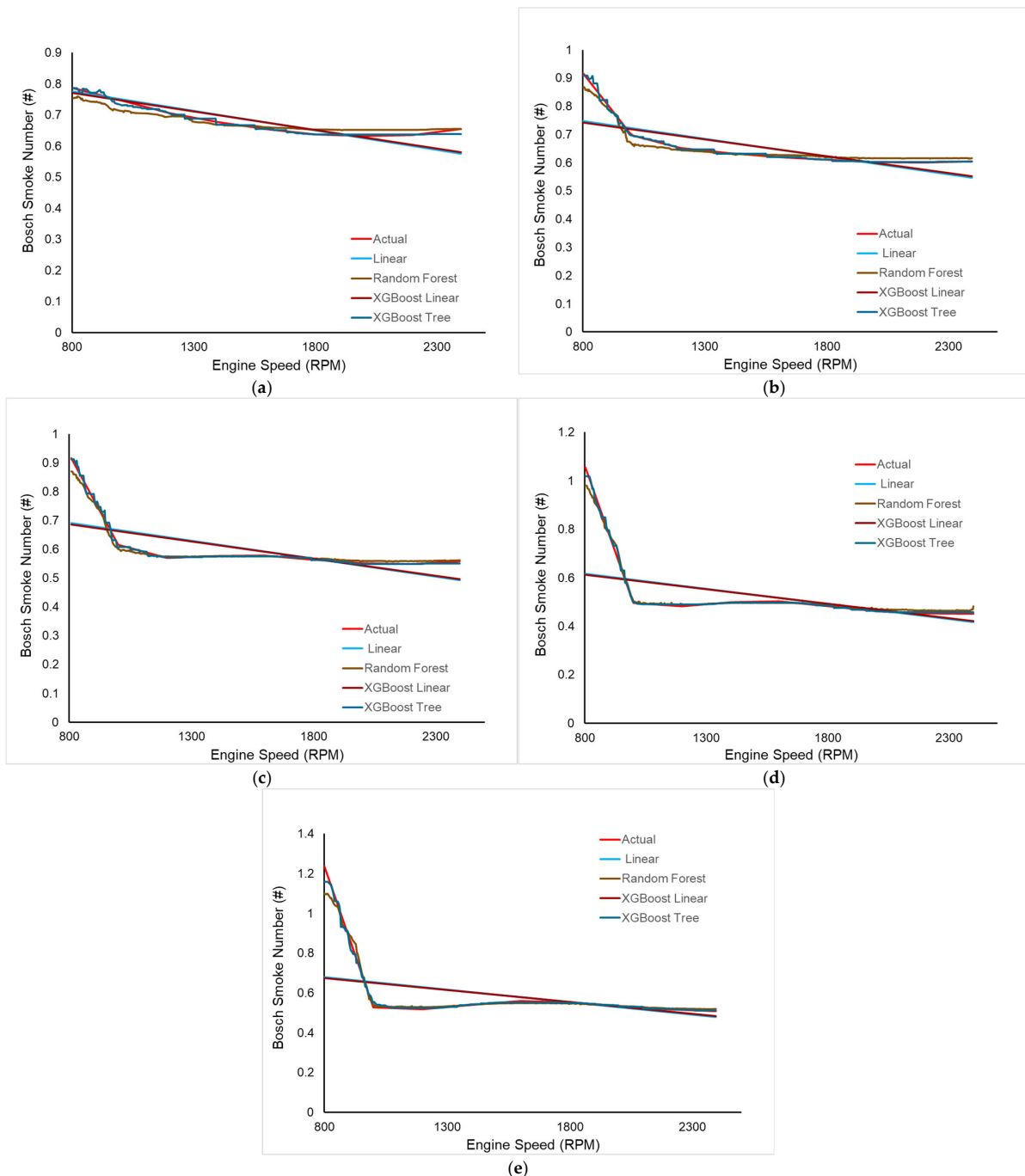


Figure 13. Smoke emission prediction using machine learning models for (a) 100D, (b) 20SME, (c) 40SME, (d) 60SME, and (e) 80SME.

The accurate prediction of the Bosch smoke number indicates that the proposed surrogate model can effectively estimate particulate-related emissions using only the engine operating conditions within the investigated operating range. Since smoke emissions are strongly influenced by the combustion characteristics of biodiesel blends, the close agreement between the predicted and simulated values demonstrates the capability of

the XGBoost Tree Ensemble model to capture the nonlinear behaviour associated with biodiesel combustion.

Overall, the machine learning results presented in Figures 9–13 demonstrate that the XGBoost Tree Ensemble Regression model consistently achieved the highest prediction accuracy across all investigated engine performance and emission parameters. The model produced the highest coefficient of determination (R^2), together with the lowest prediction errors (MAE, MSE, RMSE, and MAPE) and negligible systematic prediction bias (MSD), indicating excellent agreement with the Diesel-RK simulation results. Its superior performance can be attributed to its ability to capture the complex nonlinear relationships between engine operating conditions and the corresponding engine performance and emission characteristics. Unlike Linear Regression, which assumes linear relationships, the tree-based boosting algorithm sequentially corrects prediction errors and effectively models nonlinear interactions. Although the Random Forest Regression model also demonstrated strong predictive capability, it generally produced slightly larger prediction errors, indicating that the boosting strategy adopted by XGBoost was better suited to the simulation-generated dataset. In contrast, the Linear Regression and XGBoost Linear Ensemble models provided satisfactory predictions for relatively linear response variables, such as NO_x emissions and Bosch smoke number, but showed comparatively lower accuracy for the more complex nonlinear relationships associated with BSFC, BP, and BTE.

From a practical perspective, the developed surrogate model substantially reduces the computational effort associated with repeated Diesel-RK simulations. Once trained, the model can rapidly estimate engine performance and emission characteristics for operating conditions within the investigated domain, making it a useful tool for preliminary engine optimisation, biodiesel blend assessment, and sensitivity analysis.

These results demonstrate that the XGBoost Tree Ensemble Regression model is the most suitable surrogate model for rapidly estimating engine performance and emission characteristics within the investigated operating range. Although the model achieved very high prediction accuracy, including R^2 values approaching 1.0 for several target variables, these results should be interpreted in the context of the training dataset. The machine learning models were developed using structured simulation data generated by the validated Diesel-RK model, which inherently contains less noise and variability than experimental datasets. Consequently, higher prediction accuracy is expected compared with models trained solely on experimental measurements.

Potential overfitting was mitigated by separating the dataset into independent training (70%) and testing (30%) subsets and by applying three-fold cross-validation during model development. Furthermore, the consistent prediction accuracy observed across the training and testing datasets, together with the low prediction errors (MAE, RMSE, and MAPE), suggests that the XGBoost Tree Ensemble model learned the underlying relationships rather than memorising individual data points. Nevertheless, because the surrogate model was trained exclusively using Diesel-RK simulation outputs, its predictive capability is primarily limited to interpolation within the investigated combinations of engine speed, biodiesel blend ratio, and injection timing.

Therefore, the proposed model should not be considered a universal predictor for unseen engine configurations, alternative fuels, or operating conditions beyond those represented in the training dataset. Future work will focus on extending the dataset with additional operating conditions and experimental engine measurements to further evaluate the robustness and generalisation capability of the proposed surrogate model.

Figure 14a shows the energy and exergy rates of different SME-blended fuels with 100D. As seen in Figure 14a, the energy and exergy rates increase with increasing engine rotational speed and reach a maximum at 2400 rpm. As shown in the figure, the exergy

rate is consistently higher than that of the fuel energy rate. This is due to the fuel exergy factor, which is higher than unity. Although there are no noticeable differences in energy and exergy rates among the tested fuels, a close look at the figure reveals a slight increase in energy and exergy rates for the 20SME blended fuel compared to other SME blends. On the other hand, exhaust energy and exergy rates show the same trends as energy and exergy rates. However, SME blends exhibit lower energy and exergy rates than 80SME blended fuels, as shown in Figure 14b. When comparing the exergy efficiency of SME blends with 100D, insignificant variations in exergy efficiency were observed between the tested fuels (Figure 14c).

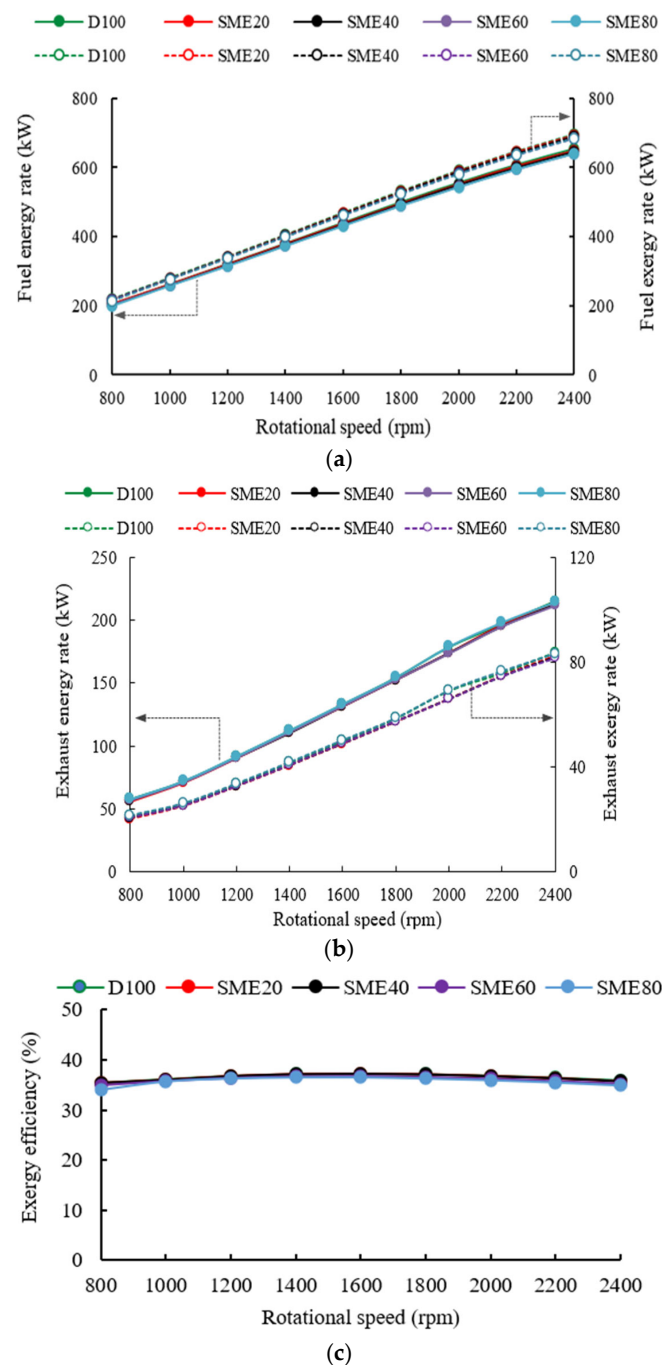


Figure 14. (a) Fuel energy and exergy rates; (b) exhaust energy and exergy rates; (c) exergy efficiency at different rotational speeds.

3.3. Diesel-RK Data Validation with GT-Suite

The investigation was further extended to validate the Diesel-RK simulation data against the GT-Suite simulation data for 1-D modelling with GT-Suite. Figure 15a,b show the GT-Suite model engine environment and the input and output parameters. The same engine configuration, including bore, stroke, compression ratio, and injection timing, was used as for the Diesel-RK software.

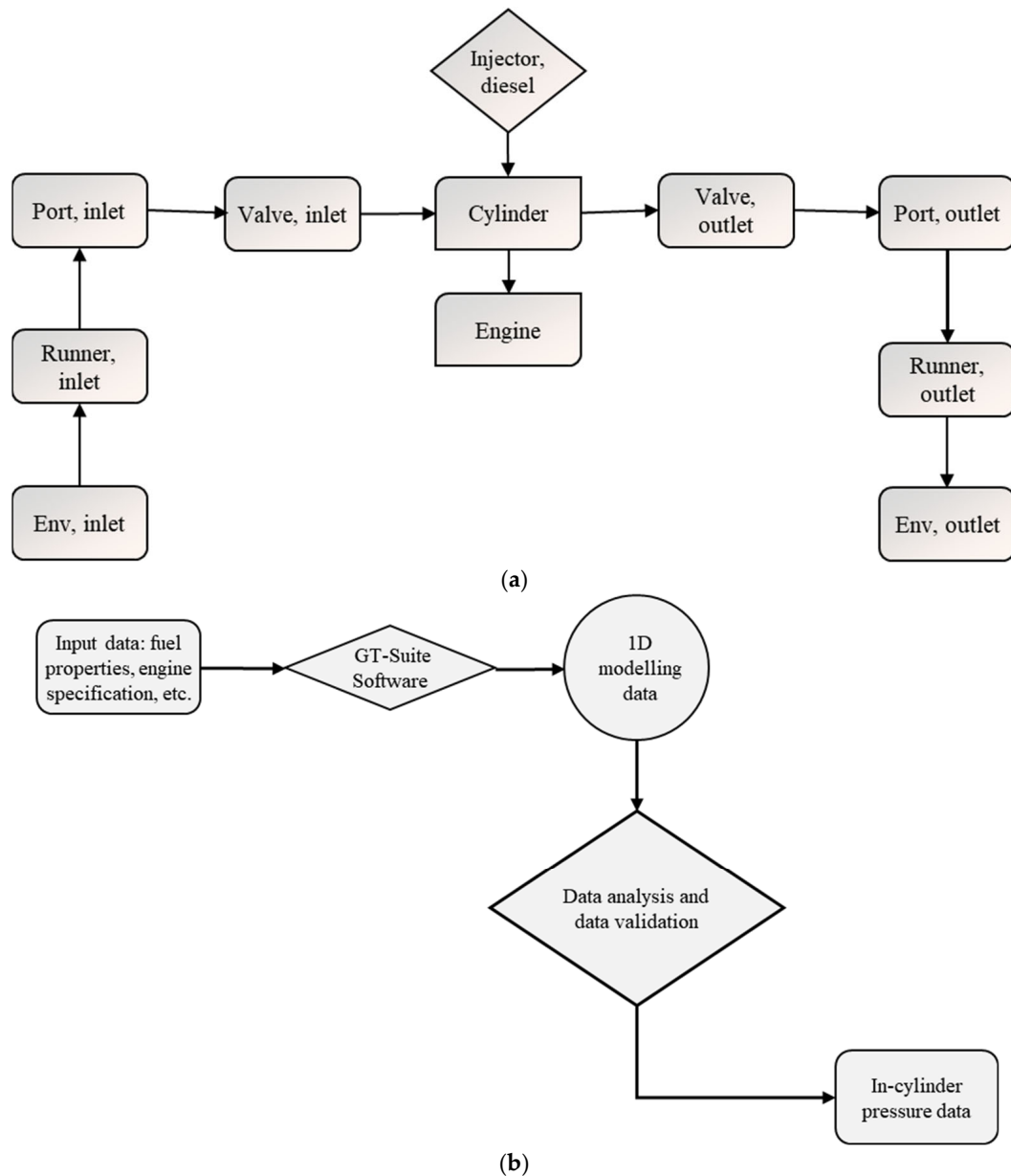


Figure 15. 1D model diagrams: (a) GT-Suite engine environment [36]. (b) Input and output parameters.

Figure 16 shows the comparison of cylinder pressure data obtained from 1D modelling from Diesel-RK against 1D modelling from GT-Suite. The cylinder pressure data was validated across four different engine rotational speeds ranging from 1400 rpm to 2600 rpm, with increments of 400 rpm. The fuel injection timing was set at 13° bTDC, and the compression ratio was kept at 22.6. A closer examination of Figure 16a–d reveals no noticeable differences in the peak cylinder pressure values predicted by Diesel-RK and GT-Suite. The maximum variation in the pressure peak was 3.7% at 1400 rpm, followed by 2.0% at 1800 rpm, 0.4% at 2200 rpm and 2.99% at 2600 rpm. Most importantly, the pressure

peaks for both Diesel-RK and GT-Suite occurred at nearly identical crank angles, indicating strong combustion phasing. These relatively acceptable deviations indicate a high level of consistency between the Diesel-RK and GT-Suite approaches, confirming the accuracy, reliability, and robustness of the developed 1D simulation model. Additionally, the close agreement in both magnitude and phasing of the cylinder pressure data demonstrates that the model accurately captures combustion behaviour across a wide range of engine operating conditions and fuels.

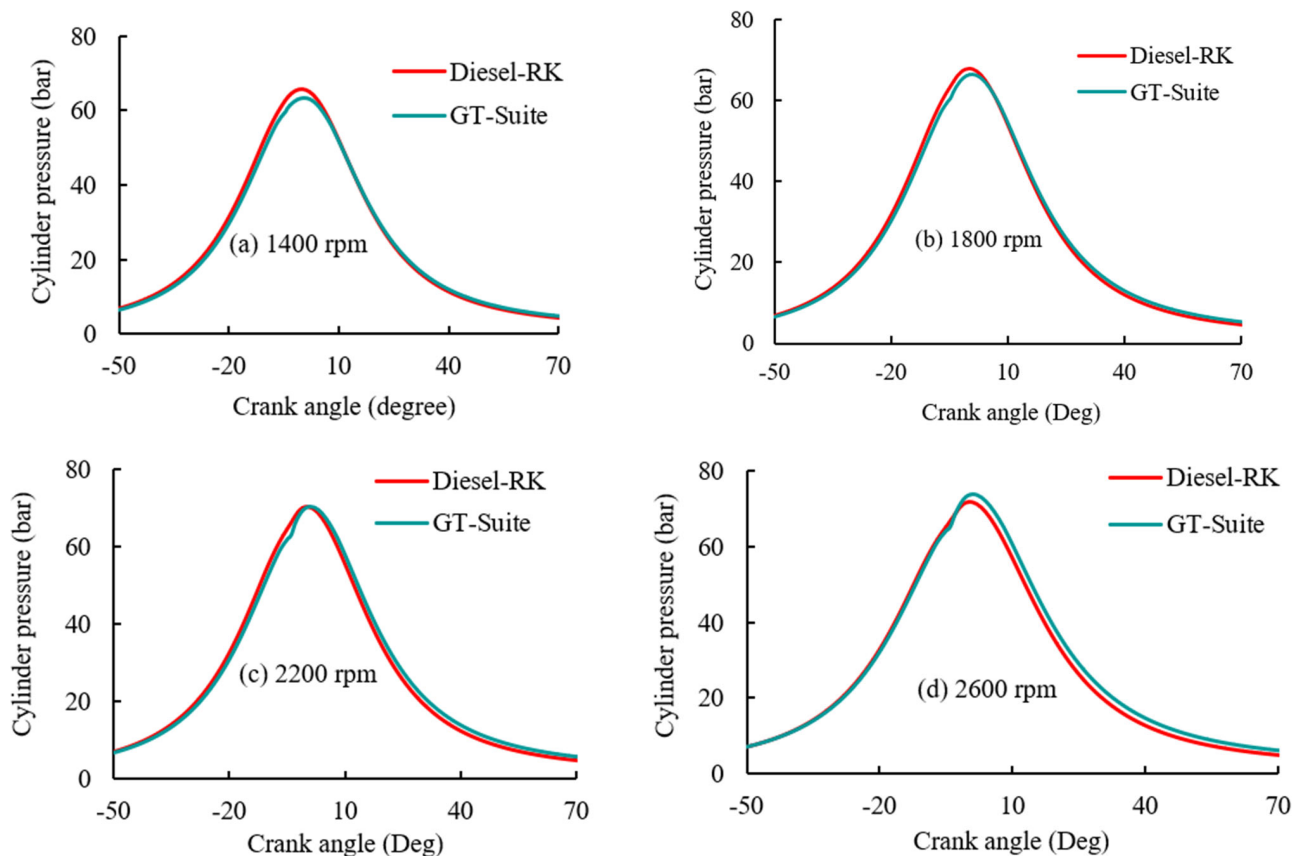


Figure 16. Cylinder pressure data for four different engine rotational speeds: (a) 1400 rpm, (b) 1800 rpm, (c) 2200 rpm, and (d) 2600 rpm. Compression ratio: 22.6; fuel injection timing: 13° bTDC. A 1D model was developed for a naturally aspirated engine.

3.4. Simulation Data Validation Against Experimental Data

In the present study, the accuracy and reliability of the model were evaluated by validating the simulated results against experimental data. For this purpose, a four-stroke, naturally aspirated, direct-injection diesel engine with a bore of 98 mm, stroke of 110 mm, compression ratio of 22.6, fuel injection timing of 13° bTDC, and engine speed of 1400 rpm was selected. The experimentally measured in-cylinder pressure data obtained under the same engine specifications and operating conditions were used as input for the GT-Suite 1D model, and the predicted results are presented in Figure 17. As illustrated in the figure, the simulated in-cylinder pressure profile closely matches the experimental measurements, demonstrating the model's ability to accurately capture the combustion characteristics. The strong agreement between the two datasets validates the developed model and confirms its suitability for predicting in-cylinder pressure under the operating conditions investigated.

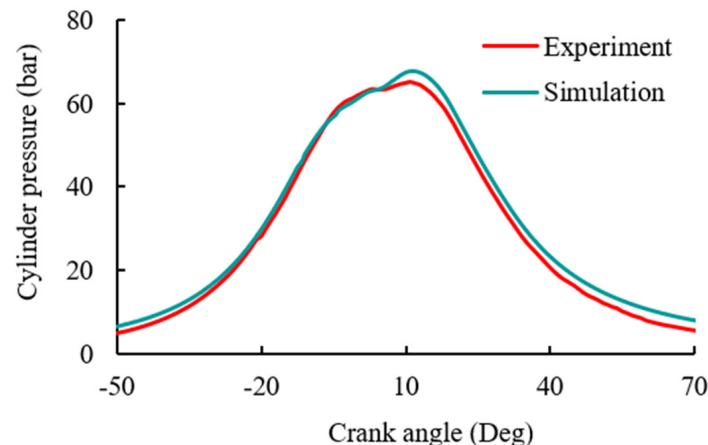


Figure 17. Validation of simulated cylinder pressure with experimental cylinder pressure data. Engine operating conditions: naturally aspirated engine; rpm: 1400; compression ratio: 22.6; fuel injection timing: 13° bTDC.

4. Conclusions

In this study, the performance, combustion, and emission characteristics of SME-blended diesel fuels in a diesel engine were investigated using a combined Diesel-RK simulation and machine learning-based prediction framework. The Diesel-RK simulation results were validated against a GT-Suite model and experimental data to verify the reliability of the proposed approach. The key findings of this study are summarised as follows:

1. SME blends demonstrated higher BSFC and lower BTE compared to diesel.
2. Similarly to BTE, the SME blends exhibited lower BMEP compared with 100D.
3. The cylinder pressure for SME blends reached the maximum at a crank angle of about 6° and was comparable to 100D.
4. The HRR profiles of SME-blended fuels exhibited an increasing trend with crank angle, reaching peak values at approximately 12° crank angle, with comparable characteristics to 100D.
5. Insignificant variations in fundamental energy and exergy parameters between the 100D and SME blends were observed.
6. NO_x emissions were higher for SME blends than for diesel. Among the SME blends, blends with higher oxygen content showed higher NO_x emissions. In contrast, smoke emissions decreased as the oxygen content of the blends increased. This was due to the fuel-borne molecular oxygen, which promotes combustion results at a higher flame temperature and consequently increases NO_x emissions while promoting more complete oxidation of soot precursors, thereby reducing smoke emissions.
7. The proposed hybrid Diesel-RK and XGBoost Tree Ensemble framework provides an efficient surrogate prediction approach for estimating engine performance and emission characteristics of SME-blended fuel and 100D. By accurately reproducing Diesel-RK simulation outputs with substantially reduced computational effort, the proposed framework offers a practical tool for rapid biodiesel blend evaluation and engine performance optimisation.
8. 1D modelling using both Diesel-RK and GT-Suite confirms the correctness and accuracy of the data. The maximum variations in pressure peaks were observed to be 3.7% and compared to experimental results; the deviation of the pressure peaks was 3.2%.

Author Contributions: Conceptualization, M.N.N.; methodology, M.N.N. and N.I.; software, M.N.N. and N.I.; validation, Q.T.; formal analysis, M.N.N.; investigation, M.N.N., F.R. and N.I.; resources, N.A.N.; data curation, M.N.N.; writing—original draft preparation, M.N.N.; writing—review and editing, F.R., N.A.N., N.I., R.Y., B.R., I.K. and Q.T.; visualization, F.R. and N.I.; supervision, M.N.N. All authors have read and agreed to the published version of the manuscript.

Funding: The research received no external funding. Central Queensland University, Australia, supported this investigation.

Data Availability Statement: Data will be available on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

BSFC	Brake specific fuel consumption
BMEP	Brake Mean Effective Pressure
NO _x	Nitrogen oxides
CO	Carbon monoxide
CO ₂	Carbon dioxide
UBHC	Unburned hydrocarbon
PM	Particulate matter
GHG	Greenhouse gas
SME	Soybean methyl ester
BTE	Brake thermal efficiency
BP	Brake power
HRR	Heat release rate
bTDC	Before top dead centre
1D	One dimensional
100D	100% Diesel
20SME	20% SME + 80% Diesel
40SME	40% SME + 60% Diesel
60SME	60% SME + 40% Diesel
80SME	80% SME + 20% Diesel
MAE	Mean absolute error
MSE	Mean squared error
RMSE	Root mean squared error
MAPE	Mean absolute percentage error
MSD	Mean signed difference

References

1. Charles, D. The Impact of Oil Prices on Guyana's Real Exchange Rate: An AlphaFold-Decomposition Analysis. *FinTech Sustain. Innov.* **2025**, *2*, A3. [\[CrossRef\]](#)
2. Bhatia, S.K.; Bhatia, R.K.; Yang, Y.-H. An overview of microdiesel—A sustainable future source of renewable energy. *Renew. Sustain. Energy Rev.* **2017**, *79*, 1078–1090. [\[CrossRef\]](#)
3. Mosarof, M.H.; Kalam, M.A.; Masjuki, H.H.; Ashraful, A.M.; Rashed, M.; Imdadul, H.K.; Monirul, I.M. Implementation of palm biodiesel based on economic aspects, performance, emission, and wear characteristics. *Energy Convers. Manag.* **2015**, *105*, 617–629. [\[CrossRef\]](#)
4. Rashid, F.; Joardder, M.U. Future options of electricity generation for sustainable development: Trends and prospects. *Eng. Rep.* **2022**, *4*, e12508. [\[CrossRef\]](#)
5. McIntosh, S.; Nabi, M.N.; Moghaddam, L.; Brooks, P.; Ghandehari, P.S.; Erler, D. Combined pyrolysis and sulphided NiMo/Al₂O₃ catalysed hydroprocessing in a multistage strategy for the production of biofuels from milk processing waste. *Fuel* **2021**, *295*, 120602. [\[CrossRef\]](#)
6. Neto, M.B.; Carvalho, P.; Carioca, J.; Canafistula, F. Biogas/photovoltaic hybrid power system for decentralized energy supply of rural areas. *Energy Policy* **2010**, *38*, 4497–4506. [\[CrossRef\]](#)

7. Mitchell, B.J.; Zare, A.; Bodisco, T.A.; Nabi, M.N.; Hossain, F.M.; Ristovski, Z.D.; Brown, R.J. Engine blow-by with oxygenated fuels: A comparative study into cold and hot start operation. *Energy* **2017**, *140*, 612–624. [\[CrossRef\]](#)
8. Hossain, F.M.; Nabi, M.N.; Rainey, T.J.; Bodisco, T.; Rahman, M.M.; Suara, K.; Rahman, S.; Van, T.C.; Ristovski, Z.; Brown, R.J. Investigation of microalgae HTL fuel effects on diesel engine performance and exhaust emissions using surrogate fuels. *Energy Convers. Manag.* **2017**, *152*, 186–200. [\[CrossRef\]](#)
9. Hedayat, F.; Stevanovic, S.; Milic, A.; Miljevic, B.; Nabi, M.N.; Zare, A.; Bottle, S.; Brown, R.; Ristovski, Z. Influence of oxygen content of the certain types of biodiesels on particulate oxidative potential. *Sci. Total Environ.* **2016**, *545*, 381–388. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Arefin, M.A.; Nabi, M.N.; Akram, M.W.; Islam, M.T.; Chowdhury, M.W. A review on liquefied natural gas as fuels for dual fuel engines: Opportunities, challenges and responses. *Energies* **2020**, *13*, 6127. [\[CrossRef\]](#)
11. Hossain, F.M.; Nabi, M.N.; Rainey, T.J.; Bodisco, T.; Bayley, T.; Randall, D.; Ristovski, Z.; Brown, R.J. Novel biofuels derived from waste tyres and their effects on reducing oxides of nitrogen and particulate matter emissions. *J. Clean. Prod.* **2020**, *242*, 118463. [\[CrossRef\]](#)
12. Hossain, F.M.; Nabi, M.N.; Rahman, M.M.; Bari, S.; Van, T.C.; Rahman, S.A.; Rainey, T.J.; Bodisco, T.A.; Suara, K.; Ristovski, Z. Experimental investigation of diesel engine performance, combustion and emissions using a novel series of dioctyl phthalate (DOP) biofuels derived from microalgae. *Energies* **2019**, *12*, 1964. [\[CrossRef\]](#)
13. Chowdhury, M.W.; Nabi, M.N.; Arefin, M.A.; Rashid, F.; Islam, M.T.; Gudimetla, P.; Muyeen, S. Recycling slaughterhouse wastes into potential energy and hydrogen sources: An approach for the future sustainable energy. *Bioresour. Technol. Rep.* **2022**, *19*, 101133. [\[CrossRef\]](#)
14. Energy Institute. *Statistical Review of World Energy*, 72nd ed.; Energy Institute, KPMG, Kearney: London, UK, 2023.
15. Lan, X.; Tans, P.P.; Thoning, K.W. Trends in Globally-Averaged CO₂ Determined from NOAA Global Monitoring Laboratory Measurements. NOAA Global Monitoring Laboratory, National Oceanic and Atmospheric Administration, U.S. Department of Commerce. 2025. Available online: <https://gml.noaa.gov/ccgg/trends/> (accessed on 8 June 2026).
16. Chiong, M.C.; Chong, C.T.; Ng, J.-H.; Tran, M.-V.; Lam, S.S.; Valera-Medina, A.; Jaafar, M.N.M. Combustion and emission performances of coconut, palm and soybean methyl esters under reacting spray flame conditions. *J. Energy Inst.* **2019**, *92*, 1034–1044. [\[CrossRef\]](#)
17. Chand, N. Plant oils-Fuel of the future. *J. Sci. Ind. Res.* **2002**, *61*, 7–16.
18. Ma, F.; Hanna, M.A. Biodiesel production: A review. *Bioresour. Technol.* **1999**, *70*, 1–15. [\[CrossRef\]](#)
19. Celikten, I.; Koca, A.; Arslan, M.A. Comparison of performance and emissions of diesel fuel, rapeseed and soybean oil methyl esters injected at different pressures. *Renew. Energy* **2010**, *35*, 814–820. [\[CrossRef\]](#)
20. Nabi, M.N.; Minami, M.; Ogawa, H.; Miyamoto, N. *Ultra Low Emission and High Performance Diesel Combustion with Highly Oxygenated Fuel*; SAE Technical Paper Series. 2000-01-0231; SAE International: Warrendale, PA, USA, 2000.
21. Nascimento, M.A.; Lora, E.S.; Corrêa, P.S.; Andrade, R.V.; Rendon, M.A.; Venturini, O.J.; Ramirez, G.A. Biodiesel fuel in diesel micro-turbine engines: Modelling and experimental evaluation. *Energy* **2008**, *33*, 233–240. [\[CrossRef\]](#)
22. Rehman, A.; Phalke, D.R.; Pandey, R. Alternative fuel for gas turbine: Esterified jatropha oil–diesel blend. *Renew. Energy* **2011**, *36*, 2635–2640. [\[CrossRef\]](#)
23. Krishna, C. *Performance of the Capstone C30 Microturbine on Biodiesel Blends*; Brookhaven National Laboratory: Upton, NY, USA, 2007.
24. Bolszo, C.; McDonell, V. Emissions optimization of a biodiesel fired gas turbine. *Proc. Combust. Inst.* **2009**, *32*, 2949–2956. [\[CrossRef\]](#)
25. Elkelawy, M.; Bastawissi, H.A.-E.; Esmail, K.K.; Radwan, A.M.; Panchal, H.; Sadasivuni, K.K.; Suresh, M.; Israr, M. Maximization of biodiesel production from sunflower and soybean oils and prediction of diesel engine performance and emission characteristics through response surface methodology. *Fuel* **2020**, *266*, 117072. [\[CrossRef\]](#)
26. Huang, J.; Xiao, H.; Yang, X.; Guo, F.; Hu, X. Effects of methanol blending on combustion characteristics and various emissions of a diesel engine fueled with soybean biodiesel. *Fuel* **2020**, *282*, 118734. [\[CrossRef\]](#)
27. Celikten, I. An experimental investigation of the effect of the injection pressure on engine performance and exhaust emission in indirect injection diesel engines. *Appl. Therm. Eng.* **2003**, *23*, 2051–2060. [\[CrossRef\]](#)
28. Gokalp, B.; Buyukkaya, E.; Soyhan, H. Performance and emissions of a diesel tractor engine fueled with marine diesel and soybean methyl ester. *Biomass Bioenergy* **2011**, *35*, 3575–3583. [\[CrossRef\]](#)
29. Diesel-RK Simulation Tool. Available online: <http://diesel-rk.ru/Eng/index.php> (accessed on 6 June 2026).
30. Kuleshov, A. Multi-zone DI diesel spray combustion model for thermodynamic simulation of engine with PCCI and high EGR level. *SAE Int. J. Engines* **2009**, *2*, 1811–1834. [\[CrossRef\]](#)
31. Nabi, M.N.; Ray, B.; Rashid, F.; Al Hussam, W.; Muyeen, S. Parametric analysis and prediction of energy consumption of electric vehicles using machine learning. *J. Energy Storage* **2023**, *72*, 108226. [\[CrossRef\]](#)

32. Khurana, S.; Saxena, S.; Jain, S.; Dixit, A. Predictive modeling of engine emissions using machine learning: A review. *Mater. Today Proc.* **2021**, *38*, 280–284. [\[CrossRef\]](#)
33. Zbikowski, M.; Teodorczyk, A. Machine Learning for Internal Combustion Engine Optimization with Hydrogen-Blended Fuels: A Literature Review. *Energies* **2025**, *18*, 1391. [\[CrossRef\]](#)
34. Lin, H.; Basem, A.; Yang, Y.; Alkhayyat, A.; Alhumaid, S.; Zhu, P.; Alotaibi, T.S.; Jaghdam, I.H.; Alkhalaf, S.; Ali, H.E. Leveraging cascade heat recovery from a gas turbine cycle in a sustainable integrated energy system; multi-objective optimization using radial basis function-driven Ant Colony algorithm. *Int. Commun. Heat Mass Transf.* **2026**, *174*, 110962. [\[CrossRef\]](#)
35. Islam, M.T.; Rashid, F.; Arefin, M.A. Numerical analysis of the performance and NO_x emission of a diesel engine fueled with algae biofuel-diesel blends. *Energy Sources Part A Recovery Util. Environ. Eff.* **2025**, *47*, 5406–5425.
36. Hussam, W.K.; Nabi, M.N.; Rasul, M.G.; Chowdhury, A.A.; Rashid, F.; Islam, N. Innovative insights into engine performance and emissions—A numerical and experimental investigation of aliphatic hydrocarbons and renewable dimethyl ether. *Case Stud. Therm. Eng.* **2025**, *72*, 106367. [\[CrossRef\]](#)
37. *GT-Suite*, version 2026; Gamma Technologies: Westmont, IL, USA, 2025.
38. Nabi, M.N.; Hustad, J.E.; Arefin, M.A. The influence of Fischer–Tropsch–biodiesel–diesel blends on energy and exergy parameters in a six-cylinder turbocharged diesel engine. *Energy Rep.* **2020**, *6*, 832–840. [\[CrossRef\]](#)
39. Demirbaş, A. Biodiesel fuels from vegetable oils via catalytic and non-catalytic supercritical alcohol transesterifications and other methods: A survey. *Energy Convers. Manag.* **2003**, *44*, 2093–2109. [\[CrossRef\]](#)
40. Graboski, M.S.; McCormick, R.L. Combustion of fat and vegetable oil derived fuels in diesel engines. *Prog. Energy Combust. Sci.* **1998**, *24*, 125–164. [\[CrossRef\]](#)
41. Lapuerta, M.; Armas, O.; Rodriguez-Fernandez, J. Effect of biodiesel fuels on diesel engine emissions. *Prog. Energy Combust. Sci.* **2008**, *34*, 198–223. [\[CrossRef\]](#)
42. Zheng, M.; Mulenga, M.C.; Reader, G.T.; Wang, M.; Ting, D.S.; Tjong, J. Biodiesel engine performance and emissions in low temperature combustion. *Fuel* **2008**, *87*, 714–722. [\[CrossRef\]](#)
43. Ramadhas, A.S.; Jayaraj, S.; Muraleedharan, C. Use of vegetable oils as IC engine fuels—A review. *Renew. Energy* **2004**, *29*, 727–742. [\[CrossRef\]](#)
44. Raheman, H.; Phadatare, A. Diesel engine emissions and performance from blends of karanja methyl ester and diesel. *Biomass Bioenergy* **2004**, *27*, 393–397. [\[CrossRef\]](#)
45. Heywood, J.B. *Internal Combustion Engine Fundamentals*; McGraw-Hill: New York, NY, USA, 1988.
46. Buyukkaya, E. Effects of biodiesel on a DI diesel engine performance, emission and combustion characteristics. *Fuel* **2010**, *89*, 3099–3105. [\[CrossRef\]](#)
47. Nabi, M.N.; Hussam, W.K.; Islam, M.J.; Rasul, M.G.; Chowdhury, A.A.; Brooks, P. Improvement of engine performance, combustion and remarkable reductions in NO_x, CO and unburned HC emissions fuelling with mixed plastic pyrolysis oil and oxygenated additives. *Case Stud. Chem. Environ. Eng.* **2025**, *11*, 101203. [\[CrossRef\]](#)
48. Nabi, M.N.; Rasul, M.G.; Hazrat, M.A.; Rashid, F. Investigation of the effect of ether oxygenated additives on diesel engine performance, combustion, and exhaust emissions—An experimental approach. *Energy* **2024**, *304*, 132223. [\[CrossRef\]](#)
49. Islam, M.A.; Rahman, M.M.; Heimann, K.; Nabi, M.N.; Ristovski, Z.D.; Dowell, A.; Thomas, G.; Feng, B.; von Alvensleben, N.; Brown, R.J. Combustion analysis of microalgae methyl ester in a common rail direct injection diesel engine. *Fuel* **2015**, *143*, 351–360. [\[CrossRef\]](#)
50. Zhang, Q.; Xu, Z.; Liu, S.; Liu, L. Effects of Injector Spray Angle on Performance of an Opposed-Piston Free-Piston Engine. *Energies* **2020**, *13*, 3735. [\[CrossRef\]](#)
51. Nabi, M.N.; Rasul, M.G.; Rahman, S.M.A.; Dowell, A.; Ristovski, Z.D.; Brown, R.J. Study of performance, combustion and emission characteristics of a common rail diesel engine with tea tree oil-diglyme blends. *Energy* **2019**, *180*, 216–228. [\[CrossRef\]](#)
52. Nabi, M.N.; Rahman, S.M.A.; Bodisco, T.A.; Rasul, M.G.; Ristovski, Z.D.; Brown, R.J. Assessment of the use of a novel series of oxygenated fuels for a turbocharged diesel engine. *J. Clean. Prod.* **2019**, *217*, 549–558. [\[CrossRef\]](#)
53. KNime Analytical Tool. Available online: <https://www.knime.com/knime-analytics-platform> (accessed on 6 June 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.