RSC Advances

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2025, **15**, 41621

Comparative analysis of the performance and emissions of a multi-cylinder diesel engine using biodiesel from underutilized feedstocks

Ashish Dewangan,*[a] Ashok Kumar Yadav,*[b] Dharamveer Singh,[c] Hasan Sh. Majdi,[d] Ali A. Rajhi,[e] Sagr Alamri[e] and Md Kareemullah *[f]

In this work, an RSM-based DOE approach was adopted to investigate and perform a comparative study on the performance and emission characteristics of a multi-cylinder transportation-type diesel engine running at variable speeds using 20% biodiesel blends of *Manilkara zapota* (MZME), *Bauhinia variegata* (BVME), Karanja (KME), and *Simarouba* (SME). The powerful desirability-based optimization technique was used to optimize system performance while reducing emissions, thereby meeting the demands of sustainable energy targets and stricter environmental regulations. To further validate the effectiveness of RSM, the results were compared with the performance of several advanced machine learning (ML) algorithms, including linear regression, decision tree, and random forest. It was found that the average reduction in BP and BTE for the blends of MZME, BVME, KME and SME was approximately in the range of 3–6%, 7–11%, 5–9% and 8–12%, respectively, whereas the increase in BSFC was 2%, 4%, 3% and 5%, respectively, compared with the diesel fuel. The average reduction in the emissions of HC and CO was in the range of 30%, 15–22%, 20–27% and 7–15% for MZME, BVME, KME and SME blends, respectively, and for NO_x emissions, the average increase was found to be 6%, 17%, 10% and 20%, respectively, compared with the diesel fuel. The random forest model demonstrated a lower MAPE than the other models, confirming its superiority in predictive modeling, with respect to both generalization and accuracy balance. The outcomes also demonstrated that both RSM and ML were highly effective modelling tools, offering accurate predictions of biodiesel performance and emission behaviour. The integrated experimental and predictive approach provided a robust framework for optimizing biodiesel formulations by identifying the optimal biodiesel blend ratio.

Received 18th June 2025
Accepted 6th October 2025

DOI: 10.1039/d5ra04352b

rsc.li/rsc-advances

1 Introduction

The growth and development of a nation are governed by its energy-producing capacity. A proportion of this requirement is fulfilled by fossil fuel sources. The modern man's complete reliance on conventional fuel for transportation has made living without petroleum extremely difficult. Mali *et al.*[1] stated that the global oil reserves are expected to run out in the coming four decades. Shahid and Jamal[2] mentioned that the transportation fuel industry contributes to 54% of the total liquid fuel utilization. The other major problem is that the excess utilization of fossil fuels has a harmful impact on nature. It adversely affects the environment, ecology and human health. The burning of diesel fuel in vehicles leads to the formation of exhaust emissions, such as carbon monoxide (CO), unburned hydrocarbons (HCs), oxides of nitrogen (NO_x), lead, sulfur oxide (SO_x), and particulate matters, which cause global climate change, an increase in greenhouse gas (GHG) emissions, photo-chemical smog, acid rain, ozone layer depletion, eutrophication and deforestation, as reported by El-Seesy *et al.*[3] The pollutants from diesel engines cause damage to crops, trees, and other vegetation by contributing to the creation of ground-level ozone. Furthermore, these pollutants cause acid rain, which affects the lakes, streams, and soil.[4] They also find their way into the human food chain through meat, fish, and water. When coal is burned, it releases a number of airborne toxins and pollutants. They include mercury, lead, sulfur dioxide, nitrogen oxides,

[a]*Department of Mechanical Engineering, Galgotias College of Engineering and Technology, Greater Noida, India. E-mail: ashishdewangan0515@gmail.com*
[b]*Department of Mechanical Engineering, Symbiosis Institute of Technology Pune, Symbiosis International (Deemed University), Pune, India. E-mail: ashokme015@gmail.com*
[c]*Mechanical Engineering Department, R. D. Engineering College, Ghaziabad, Uttar Pradesh, India. E-mail: veerdharam76@gmail.com*
[d]*Department of Chemical Engineering and Petroleum Industries, Al-Mustaqbal University, Babylon, 51001, Iraq. E-mail: dr.hasanshker@uomus.edu.iq*
[e]*Department of Mechanical Engineering, College of Engineering, King Khalid University, Abha, 61421, Saudi Arabia. E-mail: arajhi@kku.edu.sa; salamri@kku.edu.sa*
[f]*Department of Mechanical Engineering, Graphic Era (Deemed to be University), Clement Town, Dehradun, Uttarakhand, 248002, India. E-mail: mdkareemullah25@gmail.com*

particulates, and various other heavy metals. Health impacts can range from asthma and breathing difficulties to brain damage, heart problems, cancer, neurological disorders, and premature death. The combustion of natural gas produces nitrogen oxides (NO_x), which are precursors to smog, and small amounts of sulfur, mercury, and particulates. EL-Seesy *et al.*[5] mentioned that there is a pressing need for an alternative source of energy that is economically feasible, environmentally friendly and easily accessible. Alternative energy can be produced from renewable sources, which are the best choice for addressing the world's energy needs because of their vast availability and great potential. Jayashri N. *et al.*[6] reported that biofuels can be a potential alternative solution that decreases the overall dependency on petroleum-based fuels while still being able to be utilized near production sites. Ranganathan *et al.*[7] and Verma *et al.*[8] mentioned that biodiesel is a promising replacement for diesel fuel in the transportation sector because it can be used in the current diesel engines without requiring any modifications. It is a mixture of mono-alkyl esters derived from vegetable oils. Biodiesel fuel has lots of advantages as it is a sustainable, non-toxic, oxygenated, biodegradable, and sulfur-free fuel with a higher cetane rating. In addition, the emissions of HC, CO and smoke decrease when using biodiesel in diesel engines, while there is quite an increase in NO_x emissions, as also mentioned by Chattopadhyay and Sen,[9] who conducted experiments on a four-stroke, water-cooled, single-cylinder, Kirloskar diesel engine with a 3.68 kW power at 1500 rpm using biodiesel produced from cottonseed oil.

Biodiesel is widely obtained from the feedstocks of vegetable oils, animal fats, algae and microalgae. Currently, the higher cost of biodiesel generated from edible and non-edible oils is one of the primary barriers to its commercialization. The other concern is the cost of feedstock, which makes up 70% of the overall cost of biodiesel production, as was stated by Kjärstad and Johnsson.[10]

The scientific pursuit of biodiesel as a viable alternative for compression ignition engines has evolved from foundational studies establishing fundamental performance characteristics to contemporary research focused on sustainable feedstock diversification. The pioneering works by Ramadhas *et al.*[11] and Van Gerpen[12] established the inherent benefits of biodiesel, including significant reductions in particulate matter (PM), unburned hydrocarbons (HC), and carbon monoxide (CO) emissions due to the oxygen content, while simultaneously identifying the persistent challenge of increased nitrogen oxide (NO_x) emissions, creating the well-documented biodiesel NO_x–PM trade-off. The theoretical foundation linking the molecular structure to fuel performance was fundamentally established by Knothe,[13] who elucidated the structure–property relationships governing fatty acid methyl ester (FAME) compositions and their influence on combustion characteristics. Comprehensive reviews by Atabani *et al.*[14] and Mahmudul *et al.*[15] have consolidated the understanding that biodiesel properties and engine performance are intrinsically linked to feedstock fatty acid profiles, with significant variations observed across different oil sources. As sustainability concerns intensified, the field witnessed a paradigmatic shift toward nonedible and underutilized 'second-generation' feedstocks to mitigate the food-versus-fuel conflict, as highlighted by Agarwal[16] and Bhuiya *et al.*[17] The recent systematic analysis by Mofijur *et al.*[18] (2019) has demonstrated that underutilized feedstocks can achieve comparable or superior engine performance metrics to conventional edible oil-based biodiesels while addressing sustainability concerns. The contemporary reviews by Mitra *et al.*[19] and Tamilselvan *et al.*[20] have emphasized the environmental and socio-economic potential of nonedible feedstocks, particularly their relevance to national biofuel policies and regional energy security. The critical importance of process parameter optimization for maximizing underutilized feedstock potential was identified by Verma and Sharma,[21] while Demirbas[22] provided essential economic and policy contexts for driving alternative biodiesel source development. Building on these foundational and contemporary studies, the present research addresses critical knowledge gaps in the comparative performance evaluation of specific underutilized feedstocks in multi-cylinder diesel engines, contributing to the growing body of evidence supporting biodiesel resource base diversification for enhanced energy security and environmental sustainability.

In many literature studies,[23–25] various oil-bearing seeds have been mentioned, including Jatropha,[26] mahua,[27] jojoba,[28] castor,[29] linseed,[30] *Manilkara zapota*,[31] *Pongamia pinnata*,[19] and *Simarouba*.[32] Some underutilized feedstocks, such as the seeds of edible fruits like *Manilkara zapota* and nonedible seeds such as *Bauhinia variegata*, Karanja and *Simarouba*, are often considered one of the potential sources of biodiesel feedstock. *M. zapota* seeds contain 25–30% oil. Kumar *et al.*[33] reported the maximum yield of 94.83% of *M. zapota* biodiesel under the optimum conditions of a molar ratio of 6 : 1 (methanol to oil), a catalyst KOH concentration of 1% (w/w), a reaction time of 90 min and a temperature of 50 °C. Kumar *et al.*[31] conducted performance, combustion and emission analyses in an unmodified diesel engine running with biodiesel from *M. zapota* (B25 and B50) at a rated speed of 1500 rpm and compared it with diesel fuel. The B50 blends exhibited 17% higher brake thermal efficiency (BTE), 14.34% lower brake specific fuel consumption (BSFC), and 34.21% and 4.32% lower CO and HC emissions, respectively, than the diesel fuel. Agarwal and Dhar[34] studied the effect of various blends of Karanja biodiesel on the performance, combustion and emission characteristics of a DI–CI engine. The BTE and BSFC of Karanja biodiesel were found to be lower than those of the diesel fuel. The emissions of HC and smoke opacity of Karanja biodiesel were reported to be lower than those of diesel, while its emissions of CO and NO_x were found to be higher than those of the diesel fuel. Chauhan *et al.*[35] also reported a lower BTE by 3–5% of Karanja oil blends compared to neat diesel. The complete combustion of fuel results in lower emissions of HC, CO and smoke compared to the diesel fuel, but it results in an increase in NO_x emissions. Mishra *et al.*[36] investigated the effect of *Simarouba* oil methyl ester at blending ratios of 10% and 20% with the diesel fuel on a 3.5 kW DI–CI engine. At 100% load, higher BSFCs by 1.81% and 3.51%, lower BTEs by 1.90% and 3.14% and higher EGTs by 2.76% and 3.82% were reported for blending ratios of 10% and 20%, respectively, compared to the

diesel fuel. The emission analysis indicated 8–12% lower CO, 26.6–33.4% lower HC, 4.5–7.5% higher CO_2 and 4.2–8.3% higher NO_x emissions for both blends, respectively, compared to diesel.

Performance and emission studies on a DI diesel engine were performed by Kandasamy and Sundararaj[37] with *Simarouba* biodiesel. The maximum BTE was reported for the B20 blend of biodiesel. A higher BSFC of 0.3675 kg kW^{-1} h^{-1} and a lower BTE of 47.5% were observed for the B20 blend than for the diesel fuel. CO, HC, CO_2 and smoke emissions, which were 0.038%, 16.6 ppm, 3.44% and 13.26% lower than those of diesel, respectively, were reported. NO_x emission was difficult to control, which was slightly higher (1225 ppm) than that from the diesel fuel. In recent research work, an effort has also been made to use pyrolytic oil and gas fuel to reduce the dependency on fossil fuels.[38] Wu *et al.*[39] used waste tyre pyrolysis oil, which was combined with hydrogen gas, for fuel quality improvement. The addition of hydrogen resulted in an improvement in combustion, in addition to the reduction in CO, CO_2 and NO_x. Shirneshan *et al.*[40] developed an RSM-based mathematical model with the inputs of the waste cooking oil (WCO) biodiesel blend, engine load and speed to estimate the engine performance, such as brake power (BP), brake torque (BT) and BSFC. They found a decrease in BP and BT and an increase in BSFC with an increase in the blending percentage of biodiesel. Javad *et al.*[41] explored the impact of using a mixture of diesel, biodiesel, hydrogen, and Al_2O_3 nanoparticles on diesel engine performance and emissions. They found that increasing the biodiesel proportion by 30% resulted in a significant decrease in HC and CO emissions. Ali *et al.*[42] examined the influence of biofuel and lubricating oil on the concentrations, size distributions, and median diameters of PN and PM. They found that as the engine temperature increased, the PN emissions rose for all fuel types, while the average particle size decreased. Alireza *et al.*[43] evaluated the influence of biodiesel blends and alcohol on the performance and emission features of single-cylinder diesel engines using RSM. They found that adding alcohol into the mixture would reduce the BP by about 20%.

The selection of a B20 blend (20% biodiesel, 80% petrodiesel) for this investigation is rooted in its technical and commercial relevance. B20 is widely regarded as an optimal 'drop-in' fuel, offering significant emission reduction benefits without necessitating major engine hardware modifications.[31] It is supported by international standards such as ASTM D7467 and is approved for use by numerous heavy-duty engine manufacturers. While India's current mandate is B5, B20 represents a practical, next-step blending target and is the focus of numerous fleet demonstrations aimed at achieving the nation's long-term biofuel objectives.

In alignment with the global sustainability trends and national policy, this study focuses on four promising non-edible feedstocks: *Manilkara zapota*, *Bauhinia variegata*, Karanja (*Pongamia pinnata*), and *Simarouba*. The reliance on such third-generation biomass is a direct response to the critical 'food *vs.* fuel' dilemma associated with first-generation biodiesels.[19] This approach is strongly supported by India's National Policy on Biofuels, which incentivizes the use of non-edible oilseeds cultivated on non-arable land to ensure national energy security without compromising food production. While Karanja is an established candidate, the investigation into less-characterized feedstocks like *Manilkara* and *Bauhinia* is vital for diversifying the feedstock portfolio and identifying robust, locally adapted resources for sustainable biofuel production.

In most of the literature studies, the studies were carried out under various loading conditions, but at a single engine speed. In the current investigation, speeds are kept in the range of 1000–4000 rpm. The feedstocks are chosen for the study because they are underutilized and limited research work has been carried out on their engine performance and emissions. The novelty of the study lies in the comparative study of the selected feedstock in a multi-cylinder engine under various speed conditions. The physicochemical characterizations of biodiesels from all the feedstocks have been compared. Further, the experimental analysis is performed on the diesel engine using all four biodiesel fuels at a blending ratio of 20%, and its performance (BP, BTE and BSFC) and emission (HC, CO and NO_x) characteristics have been evaluated and compared with the engine running on the diesel fuel at different engine speeds.

While numerous studies have investigated the use of biodiesel blends in internal combustion engines, a significant research gap remains in the systematic integration of statistical and machine learning (ML) techniques for comprehensive performance and emission optimization. Most existing research tends to emphasize either experimental investigations or the application of individual modelling methods in isolation. Experimental approaches, while valuable, are often limited by time, cost, and the number of feasible trials. By contrast, modelling studies frequently lack experimental validation or fail to account for the complex, nonlinear interactions between multiple variables such as blend ratio, injection timing, engine load, and speed. Furthermore, few studies have explored how statistical design tools like response surface methodology (RSM) can be effectively combined with machine learning algorithms to create a more accurate and generalizable predictive framework. This lack of an integrated methodology limits the ability to capture the full potential of biodiesel optimization strategies. Therefore, there is a clear need for a hybrid approach that not only models but also validates and refines predictions using a combination of experimental data, statistical optimization, and data-driven machine learning techniques.

2 Materials and methods

2.1 Materials

Karanja oil seeds and *Bauhinia variegata* seeds were procured from a local vendor in Delhi. *M. zapota* seeds were collected from the Delhi Technological University campus. *Simarouba* seeds were collected from a rural area in the Bhubaneshwar district, Orissa, India.

2.1.1 *Manilkara zapota* (MZ) seeds. The MZ fruit is also known by the name sapodilla. It is an evergreen tree and drought-tolerant. It produces berry-type fruits with a width of 4–8 cm. It is primarily cultivated for fresh consumption and is commonly served as a dessert fruit. Each fruit holds up to 6

seeds. It is grown on around 564 hectares in tropical and subtropical lowlands. It is mostly found in Asian countries. Its fruits and leaves are used for making a decoction for the treatment of ulcers, fever, indigestion, hemorrhage, and wounds. The seeds contain 23–30% oil, and they might be considered as a potential feedstock for biodiesel production.[44]

2.1.2 Karanja seeds. The Karanja tree is available in all parts of India. It can be commonly seen on roadsides and canals to stop soil erosion. It is an evergreen and deciduous tree. Its flowers are white, lavender or pink. Its pods contain 1–2 seeds. The seeds contain 30–40% oil. They might be a good substitute for diesel.[45]

2.1.3 Bauhinia variegata seeds. Bauhinia variegata belongs to the family of Leguminosae (Caesalpinioideae). It is also known by other names such as Kachnar, mountain ebony and camel's foot tree. It is an evergreen, deciduous tree, and it can grow up to a height of 15 m. Its twigs are angled slender. Its inner bark is fibrous, while the outer one is fissured and scaly. Bauhinia variegata is spread across India, China, Myanmar, Nepal, Pakistan, and Sri Lanka. Its flowers are vitamin C-rich. It has a great use in Ayurveda, Unani, and homoeopathy medical practice. Its bark and roots are used as medicine for the treatment of scrofula, lymphadenitis, worm infestations, and skin diseases and as an antidote for snake poison. Chemical compounds such as triterpine, beflavone, flavonoids, phenanthraquinone, saponin and flavonolglycoside are found in its herbs, which are anti-inflammatory, antifungal and antibacterial.[46]

2.1.4 *Simarouba* seeds. The *Simarouba* tree belongs to the family of Simaroubaceae. It is an evergreen tree with an average height of 12–15 m. It is also called malacacheta, bitter wood, maruba, paradise tree, and simaba. It is native to Mexico, Jamaica, Cuba and the USA. Most parts of the tree are used in the field as manure, medicine and fuel. It can be used in the production of ethanol, cattle food, and the charcoal industry and for making surfactants, soaps and detergents. It has good medicinal properties, such as antibacterial, antidysenteric, sudorific, and tonic.

2.2 Methodology

Four non-edible feedstocks were selected, namely, *Manilkara zapota*, *Bauhinia variegata*, Karanja and *Simarouba*, for biodiesel production. A Soxhlet extraction system was employed for oil production from seeds. The fatty acid composition of the produced oil was determined by GC. The transesterification reaction was employed for biodiesel production from the oil. A 20% blend of biodiesel with diesel was used for performance and emission analyses in a diesel engine. All experiments on the CI engine were conducted three times under identical operating conditions to ensure the repeatability and consistency of the measured data. The values reported in this work represent the average of these repeated trials. The variation between repetitions was found to be negligible, thereby confirming the reliability of the reported results.

2.2.1 Oil extraction. The seeds of *Manilkara zapota* (L.), Karanja, *Bauhinia variegata* and *Simarouba* were kept in sunlight for moisture removal. After a few days, the dried seeds were detached from the shell and ground. Then, they were placed in the Soxhlet extraction system. For the extraction of oil, petroleum ether (solvent) was mixed with the seeds, which underwent repeated cycles of condensation and evaporation processes. The extracted oil–solvent mixture was heated at 60 °C for the separation of oil from the solvent. Sodium sulphate anhydrous was mixed with the oil and further heated for the removal of moisture. The seeds of *M. zapota* contain approximately 30–35% oil by weight. The percentage oil content in *B. variegata* obtained is 19%. *Simarouba* fruit seeds contain approximately 69% oil, while Karanja seeds contain 30% oil.

2.2.2 Determination of the fatty acid composition (FAC). The FACs of *M. zapota*, *Bauhinia variegata*, Karanja and *Simarouba* oils were determined by a GC-6890 EZ Chrome Elite gas chromatograph. Before injecting, the oil sample was prepared by mixing 67.53 g of it with 20 mL of a *trans*-methylene mixture to form fatty acid methyl esters (FAMEs). Petroleum ether and distilled water were then separated from the FAME. Further, it was reconstituted with petroleum ether and then injected into the GC. The temperature was kept at 100 °C for 2 minutes initially and increased at a heating rate of 10 °C min^{-1} to 240 °C. After keeping it for 7 min, the carrier gas (N_2) was supplied at 8 mL min^{-1}. A flame was produced by supplying hydrogen and air at 50 and 400 mL min^{-1}, respectively. The FAC of the oil sample was recognized by comparing its retention time with the standard FAME-37 mix. Table 1 shows the FACs of the selected oils, *i.e.*, *Manilkara zapota*, *Bauhinia variegata*, Karanja and *Simarouba* oils. The unsaturated fatty acid content in the test fuels followed the increasing order of *Simarouba* oil (57.6%), *B. variegata* (63.58%), Karanja (80.4.%) and *M. zapota* (83.54%). Biodiesel is said to be stable if the amount of unsaturated fatty acid is less, which can also be its oxidation resistance indicator.

The performance and emission characteristics of the tested biodiesel blends are fundamentally influenced by their fatty acid compositional profiles, as presented in Table 1. The degree of saturation varies from 16.46% in *M. zapota* to 42.1% in *Simarouba* oil, while that of unsaturation ranges from 83.54% to 57.6%, respectively. These structural differences, particularly the balance between saturated fatty acids (affecting the cetane number and ignition quality) and unsaturated fatty acids (influencing the fuel fluidity and combustion temperature), create distinct combustion characteristics that manifest in the observed engine performance and emission trends.

2.2.3 Production of biodiesel from oils. The free fatty acid (FFA) contents of *M. zapota*, Karanja, *B. variegata* and *Simarouba* oil were determined to be 1.76%, 8.5%, 1.8% and 2.8%, respectively. Thus, a single-step transesterification (alkaline-base catalyst) process was used to convert *M. zapota*, *B. variegata* and *Simarouba* oils into biodiesel (because their FFA < 2%), whereas a two-step transesterification process (acid–base catalyst), *i.e.*, esterification, followed by a transesterification process, was used to convert Karanja oil into biodiesel (because its FFA > 2%).

2.2.3.1 Esterification process. In the esterification process, the fatty acid was converted to methyl ester, which also lowered its acid value. The 12 : 1 molar ratio was taken for methanol to

Table 1 Fatty acid composition of *M. zapota*, Karanja, *B. variegata* and *Simarouba* oils

Fatty acids	Chemical formula	Degree of unsaturation	*M. zapota* oil	Karanja oils	*B. variegata* seed oil	*Simarouba* oil
Palmitic	$C_{16}H_{32}O_2$	16 : 0	14.2	11.8	20.10	11.3
Stearic	$C_{18}H_{34}O_2$	18 : 1	2.26	7.8	15.31	30.7
Oleic	$C_{18}H_{32}O_2$	18 : 2	63.83	52.30	25.92	52.8
Linoleic	$C_{20}H_{40}O_2$	20 : 0	18.20	16.6	37.11	2.5
Linolenic acid	$C_{18}H_{32}O_2$	18 : 3	1.51	2.7	0.55	0.3
Arachidic	—	20 : 0	—	1.9	—	1.7
Eicosenoic	$C_{20}H_{40}O_2$	20 : 1	—	1.2		
Behenic acid	—	22 : 0	—	4.6		0.3
Lignoceric	—	24 : 0	—	1.1		
∑saturated fatty acids			16.46	19.60	35.41	42.1
∑unsaturated fatty acids			83.54	80.40	63.58	57.6

Table 2 Total production cost/litre of biodiesel in Indian rupees

Types of seeds	Cost of seeds	Manufacturing equipment cost (electricity, *etc.*)	Chemical cost	Labour cost	Total production cost
M. zapota	25	6	5	5	41
B. variegata	50	6	5	2	63
Karanja	45	8	6	3	62
Simarouba	42	6	5	2	55

Karanja oil. The 2% (v/v) mixture of pre-heated oil and H_2SO_4 was maintained at 500 rpm for 2 h at 60 °C. The reaction products were shifted to a separating funnel, and after a sufficient settling time, the top layer (impurities, H_2SO_4 and excess alcohol) was separated from the bottom layer. The esterified oil was then separated from water and methanol by keeping it under vacuum conditions for 1 h at 95 °C in a rotary evaporator.

2.2.3.2 Transesterification process. In the transesterification reaction (Fig. 1), biodiesel was produced by mixing the test oil with 25% (v/v) methanol and a 1% (m/m) KOH catalyst and keeping the mixture at 500 rpm for 2 h at 60 °C. The biodiesel was shifted to a separating funnel for 6–8 h to separate it from glycerol. After clearly seeing the top and bottom layers, the bottom layer, which consisted of glycerol and impurities, was removed. The biodiesel yields from *Manilkara zapota*, *Bauhinia variegata*, *Simarouba* and Karanja oils were found to be 95%, 94%, 98.1% and 97%, respectively.

2.2.4 Biodiesel–diesel blends. In this investigation, the blends of 20% biodiesel with diesel (*i.e.*, 20% biodiesel + 80% diesel) on a volume basis were prepared for the produced *M. zapota*, *B. variegata*, Karanja and *Simarouba* oils and designated as MZME20, BVME20, KME20 and SME20, respectively. The impact of 20% blend on physicochemical properties of biodiesel is observed and represented in Table 5.

2.3 Economic viability of biodiesel feedstock

2.3.1 Baseline production cost comparison. The cost of making biodiesel is determined by a number of factors, including chemical costs, manufacturing machinery costs, labour costs, and the cost of biodiesel raw materials from industrial sources. Table 2 shows the various costs associated with biodiesel production, such as electricity use. The production cost of *M. zapota* biodiesel is the lowest among all selected feedstocks.

The baseline analysis indicates that *M. zapota* is the most economical feedstock, primarily due to its significantly lower seed cost. The feedstock cost is clearly the dominant variable, accounting for 61% of the total cost for *M. zapota* and rising to 79% for *B. variegata*. However, this represents a simplified "gate-to-gate" assessment. A robust evaluation of commercial viability must consider several external factors.

2.3.2 Influence of supply chain logistics and feedstock availability. The "seed cost" in our analysis reflects the market price at the time of purchase and does not include logistical overhead. In a commercial scenario, the cost of collection and transportation would be a major factor. For instance, Karanja is widely available and often grown in semi-plantations, which could streamline collection. In contrast, *M. zapota* seeds are typically a horticultural waste product from scattered trees, which would require a more complex and potentially costly collection network. Furthermore, the oil yield of the seeds is critical; a low-cost seed with a very low oil yield may be less economical overall when processing costs per litre of the final product are considered.

2.3.3 Impact of market and seasonal fluctuations. The costs presented in Table 2 are a snapshot based on data from Q4 2024. In reality, feedstock prices are subject to significant seasonal variations, with costs typically being the lowest during the harvest season and increasing thereafter. For example, the price of Karanja seeds can fluctuate by up to 30–40% annually. Similarly, the costs of key chemical inputs like methanol and sodium hydroxide are tied to volatile global commodity

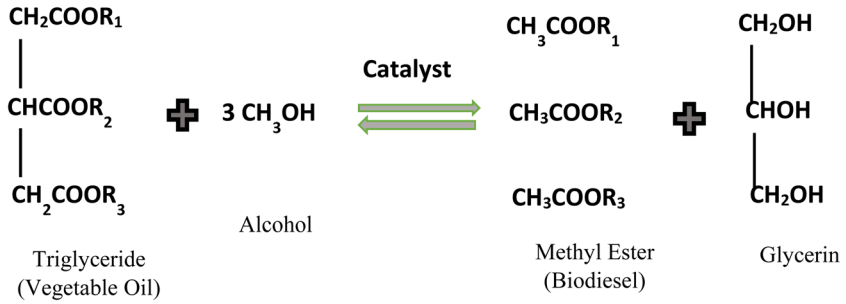


Fig. 1 Transesterification reaction.

markets. Therefore, the economic viability of any single feedstock would depend on strategic year-round sourcing and price hedging, which are beyond the scope of this analysis but are critical for industrial applications.

In conclusion, while *M. zapota* shows significant promise based on its low direct input costs, its commercial success would be contingent on developing an efficient collection supply chain. This analysis highlights that for biodiesel to be economically competitive, research and policy must focus not only on conversion technology but also on optimizing the entire feedstock supply chain.

2.4 Response surface methodology

When many independent variables affect a response, response surface methodology (RSM) is a robust statistical tool used to model and maximize processes. Analyzing and enhancing complicated systems especially benefit from it. The approach usually generates a second-order polynomial mathematical model that explains the link between the input variables and the intended result.[47] RSM uses past data to enable response prediction and optimization without requiring further experimental designs. The method starts with choosing independent variables depending on past data. These variables are methodically changed within their observed ranges to create a regression model most appropriate for the response surface. As a multi-dimensional graphical depiction, the response surface displays how variations in the inputs influence the output. Depending on the aim, the best mix of input variables is sought to maximize or decrease the response.[48]

RSM's capacity to detect interactions between variables that can greatly affect the result is a fundamental characteristic. In particular, in non-linear systems, RSM offers a more accurate depiction of how the factors affect the response by including interaction and quadratic terms in the model. This prediction ability enables scientists to modify variables to improve process performance without further investigation.[49] Multi-objective optimization is much aided by desirability. Each response, when several are involved, becomes a desirability function,

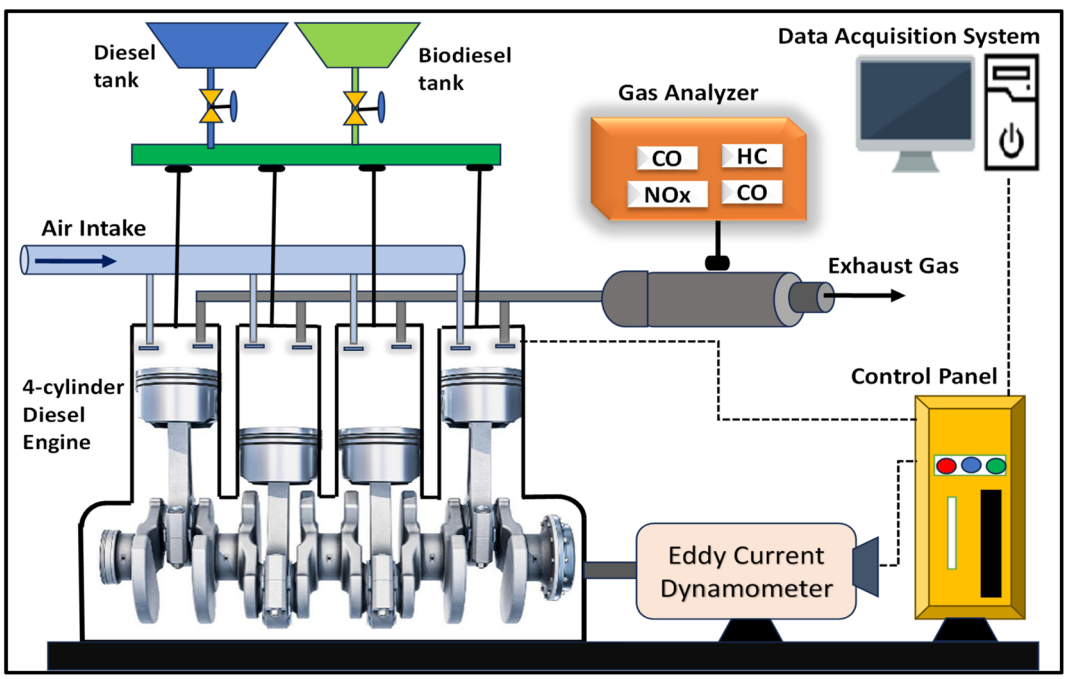


Fig. 2 Experimental set-up.[56]

scaled from 0 to 1. A desirability score of 1 denotes the most favorable outcome; a score of 0 denotes an undesired result. Combining individual desirability for every response creates an overall desirability score that lets researchers maximize every response at once. Maximizing the general desirability guarantees the best feasible result across all the goal variables in the system.[50]

2.5 Machine learning

2.5.1 Linear regression. A fundamental statistical method used to approximate the relationship between a dependent variable and one or more independent variables is linear regression. The main objective is to derive a linear equation that, given the independent variables, fairly forecasts the dependent variable. In simple linear regression, this involves developing an equation using parameters, like the intercept and slope, that link the projected value to the independent variable. The least squares approach is used to estimate the coefficients of this equation, thereby reducing the sum of the squared deviations between the observed and expected values.[51] Multiple linear regression carries this idea forward to include several independent variables, thereby enabling the evaluation of their combined influence on the dependent variable. Underlying linear regression are several presumptions: linearity of relationships, independence of data, and homoscedasticity, that is, constant variance of residuals. Different criteria, including the coefficient of determination and adjusted coefficient of determination, are applied to assess the performance of the model. Furthermore, hypothesis testing on the coefficients aids in the statistical relevance of the predictors, thereby enabling the wise selection of variable inclusion in the model.[52]

2.5.2 Random forest. In this study, Random Forest (RF) was employed to strengthen the predictive modelling of engine performance and emission features when using biodiesel blends. While single decision trees are simple and easy to interpret, they often suffer from over-fitting and may not perform well on unseen data. Random Forest overcomes this by building a collection of decision trees, each trained on a random subset of the data, and then combining their predictions to reach the final output. This ensemble approach helps balance the trade-off between bias and variance, improving the overall model accuracy and robustness.[53]

Each tree in Random Forest is trained using a technique called bootstrap sampling, where it learns from a randomly selected portion of the dataset (with replacement). Moreover, at each decision point (or node) within the tree, only a random subset of the features is considered. This intentional randomness—both in the data and feature selection—ensures that the individual trees are diverse, reducing the likelihood that the model will over-fit the training data.[54]

For regression tasks like ours—predicting brake thermal efficiency (BTE), brake specific fuel consumption (BSFC), and emissions (*e.g.*, No_x, CO, and HC)—the final prediction from Random Forest is obtained by averaging the outputs of all the trees. This averaging smooths out fluctuations caused by any single tree and produces more stable and reliable results.

One key advantage of Random Forest is its ability to identify which input variables have the most influence on the output. In this study, the algorithm helped highlight the relative importance of variables. This is especially helpful for researchers aiming to fine-tune operating conditions or improve fuel formulations.

Another practical strength of RF is that it works well even when the dataset contains noisy, missing, or redundant features. It also handles high-dimensional data without requiring extensive preprocessing or manual feature selection.

To assess the model's performance, several metrics were used, including root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2). The model consistently achieved high R^2 values (>0.95), showing that it could accurately capture the complex relationships between input variables and engine responses.

Table 3 Technical specification of the test engine

Make	Telco, Model Tata Indica
Type	4 cylinder, 4 stroke, diesel, water-cooled
Rated power	39 kW at 5000 rpm
Torque	85 N m at 2500 rpm
Cylinder volume	1405 cc
Compression ratio	22 : 1
Dynamometer	Type = eddy current, water-cooled, with a loading unit
Load sensor	Load cell, type strain gauge, range = 0–50 kg
Software	"Engine soft" engine performance analysis software

Table 4 Range, accuracy and uncertainty of measurements

Measurements	Instrument	Accuracy
Engine power	—	±1.0%
Fuel consumption	Level sensor	±2.0%
Air consumption	Turbine flow type	±1.0%
BTE	—	±1.0%
BSFC	—	±1.5%

Table 5 Technical details of the emission analyzers

Emission analyzer	Measuring range	Accuracy	Percentage uncertainties
Carbon monoxide (CO)	0–10% vol	±0.02% vol	±0.2
Carbon dioxide (CO_2)	0–20% vol	±0.03% vol	±0.15
Hydrocarbon (HC)	0–20 000 ppm	±15 ppm	±0.2
Nitric oxide (NO)	0–5000 ppm	±50 ppm	±0.2

2.5.3 Decision tree. A decision tree is like a flowchart that helps make decisions based on data. Each decision starts at the top (called the root), and as you move down, you follow different paths (branches) depending on the answers to certain questions or rules for the data. Each branch represents a decision rule, and at the end of each branch (called a leaf), you find the final outcome or prediction.

Building a decision tree begins by looking at the data and picking the feature (or characteristic) that best helps to separate the data into different categories. This is done by measuring how much each feature helps reduce uncertainty. Think of it like finding the feature that gives the clearest divide—one that leads to the most accurate predictions. Common ways to measure this include the Gini impurity or entropy,[55] both of which essentially tell us how mixed up or impure the data is after the split.

Once the best feature is chosen, the data gets split into smaller groups, and the process repeats recursively for each new group. This continues until certain rules are met—like when the tree gets too deep or when enough data points are not left in a group to split further.

While decision trees are great because they are easy to understand (you can literally follow the branches to see how decisions are made), they do have a downside. If the tree is too detailed, it can end up "memorizing" the data rather than learning the general pattern, which is called overfitting. Overfitting happens when the tree gets too complex, and it performs poorly on new, unseen data because it is too focused on the specifics of the training data.

To prevent this, one of the main techniques used is called pruning, which is just a fancy way of saying: we trim the tree by cutting out branches that are not really helping with predictions. This makes the tree simpler and better at handling new data.

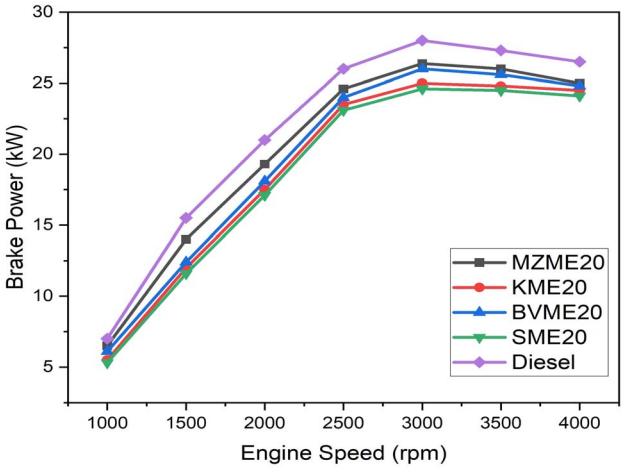


Fig. 3 Brake power variation with engine speed.

Decision trees are very flexible because they can work with both numerical and categorical data. This makes them perfect for a wide range of tasks, from predicting if someone will click on an ad (a classification problem) to forecasting the price of a stock (a regression problem). Additionally, because they are easy to understand, they are often used when it is important to know exactly why a certain decision was made.

3. Experimental setup

3.1 Test setup

The experiments were carried out in a Telco, Tata Indica multi-cylinder transportation-type diesel engine test rig. The engine was coupled with an eddy current-type dynamometer. The engine schematic diagram is depicted in Fig. 2.

Table 3 presents the technical details of the engine. A 20% blend ratio of biodiesel from all feedstocks, namely, MZME20,

Table 6 Properties of *M. zapota*, Karanja, *B. variegata* and *Simarouba* oils and their biodiesel blends w.r.t. diesel

Property	Test method	ASTM D6751 and EN 14214	MZO	KSO	BVO	SO	MZME	KME	BVME	SME	MZME 20	KME 20	BVME 20	SME 20	Diesel
Density at 15 °C (kg m^{-3})	D 1298	860–900	0.886	0.931	0.91	0.89	0.878	0.884	0.86	0.867	0.854	0.864	0.84	0.838	0.83
Viscosity at 40 °C (mm^2 s^{-1})	D 445	1.9–6.0	33.8	38.7	21.43	49.56	4.7	4.5	4.3	4.1	3.3	3.1	3.2	3.72	2.91
Calorific value (MJ kg^{-1})	D240	—	39.1	38.3	37.12	35.5	40.2	39.1	39.15	37.8	43.12	42.2	42.7	41.458	44.85
Flash point (°C)	D 93	Min 120	271	226	200	235	179	168	187	162	85.4	78	89	67	71.5
CFPP (°C)	D 6371	Max 19	—	—	6		−3	−2	4		−7	−6			−13
Pour point (°C)	D97	−15 to 16	—	−2	−2	14	−5	−4	−6		−8	−8	−7	−10	−14
Cloud point (°C)	D 2500	−3 to 12	—	2	5	17	−4	−5	3		−9	−7			−16
FFA (%)	D664	—	1.76	8.5		3.2	0.30	0.22		0.18	—	—			0.014
Acid value (mg_{KOH} g^{-1})	D 664	Max 0.5	3.52	17	2.2	6.56	0.18	0.45	0.14	0.43	—	—	—	0.32	0.01
Iodine value (mgI_2/100 g)	EN 14214	Max 120	64.95	86.2		84.5	66.15	90		81.2	—	—	—	—	—
Saponification value (mg g^{-1})	—	—	187	184	183.5	184	174	178.5	164.7	170	—	—	—	—	
Cetane number	D 613	Min 51	—	38.0	45	48.2	52.5	45	47	57.4	51.2	48	50	48.5	51.3

BVME20, KME20 and SME20, was taken as the engine fuel. Initially, the diesel fuel was used for starting and warming up the engine. Then, it was switched to biodiesel fuel, *i.e.*, the engine was then run with MZME20, and engine performance and emission characteristics were recorded. Subsequently, the engine was run with the blends of BVME20, KME20 and SME20 for the measurement of characteristics. The experiment was performed within the 1000–4000 rpm range under full-load conditions. For performance evaluation, the 'Engine soft' software was used.

An AVL Di gas analyser was connected to the engine's tail pipe for the measurement of engine exhaust emissions such as HC, CO and NO_x. A non-dispersive infrared (NDIR) sensor was employed for HC and CO emission measurement, and NO_x was measured using a photochemical sensor. The range, accuracy and uncertainty of measurements of the instruments are presented in Table 4.

Table 5 shows the technical details of emission analysers.

3.2 Uncertainty analysis

Uncertainty analysis is the quantification of the error that occurs during experimentation. The sources of errors are equipment calibration, selection, ambient conditions, *etc.* The percentages of uncertainty for the measurement of BTE, BSFC, HC, CO and NO_x are shown in Tables 4 and 5. The overall uncertainty can be calculated using eqn (1):

$$\begin{aligned}\text{Overall uncertainty (\%)} &= \sqrt{\left[\text{uncertainty of (BTE)}^2 + (\text{BSFC})^2 + (\text{CO})^2 + (\text{HC})^2 + (\text{NO}_x)^2\right]} \\ &= \sqrt{\left[(\pm1)^2 + (\pm1.5)^2 + (\pm0.2)^2 + (\pm0.2)^2 + (\pm0.2)^2\right]} \\ &= \pm1.835\end{aligned} \quad (1)$$

4. Results and discussion

4.1 Properties of selected oils and their biodiesel

Standard techniques were applied for properties determination of *Manilkara zapota*, Karanja, *Bauhinia variegata* and *Simarouba* oil biodiesel. The properties of the test fuel oils, their biodiesel and 20% blends are shown in Table 6. The CV and cetane index number of MZ biodiesel are superior to those of other biodiesels.

4.2 Performance and emission features

4.2.1 Brake power (BP). Fig. 3 depicts the change in the BP with engine speed for all the test blends of MZME20, BVME20, KME20, SME20 and neat diesel. The engine running with diesel shows a higher BP than that running with all biodiesel feedstocks. A similar trend was found by Ileri *et al.*[57] Similarly, Ali *et al.*[58] found the same trend for BP; they conducted an experiment using a four-cylinder diesel engine fuelled with palm oil biodiesel and found a BP lower by 0.5–2.7% for the B10–B30 blend ratios compared to the diesel fuel. An increase in the value of BP is observed with an increase in engine speed up to 3000 rpm for all test fuels, and above this speed, the BP decreases due to the higher friction effect. At 3000 rpm, the values of BP (in kW) are found to be 28, 26.38, 26, 25 and 24.6 for diesel, MZME20, BVME20, KME20 and SME20, respectively. The higher density and viscosity and lower CV of biodiesel are the possible reasons for its lower BP than that of neat diesel. The average reductions in the BP for MZME20, BVME20, KME20 and

Table 7 ANOVA results for engine performance

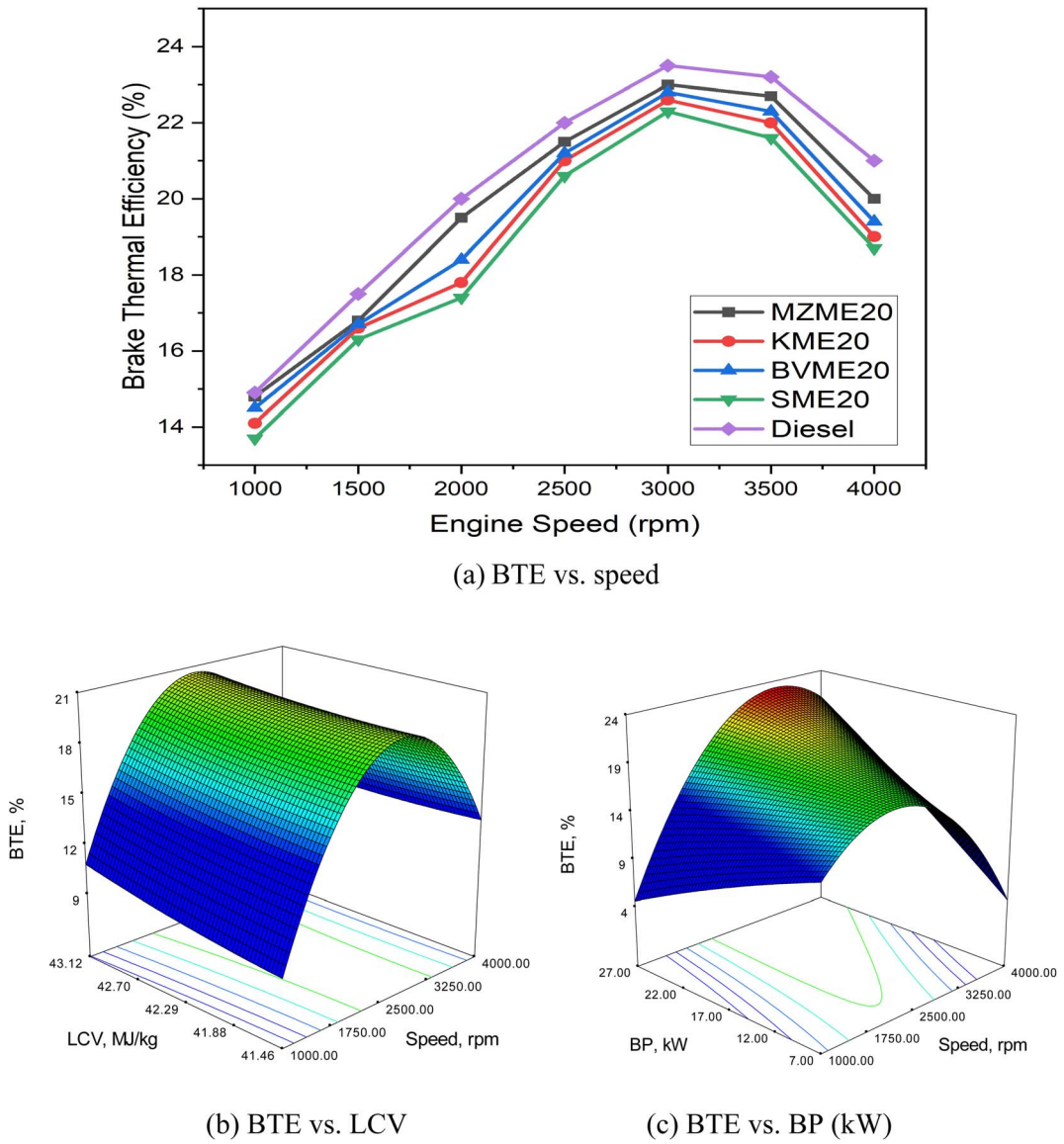
	BTE			BSFC		
Source	Sum of squares	*F* value	*p*-Value prob > *F*	Sum of squares	*F* value	*p*-Value prob > *F*
Model	224.37	91.94	<0.0001	36 289.94	146.99	<0.0001
A-speed	0.09	0.32	0.58	0.41	0.01	0.90
B-LCV	0.46	1.69	0.21	43.82	1.60	0.22
C-BP	0.07	0.26	0.62	2.24	0.08	0.78
AB	0.25	0.91	0.35	63.19	2.30	0.15
AC	0.81	2.99	0.10	10.87	0.40	0.54
BC	0.01	0.05	0.82	224.01	8.17	0.01
A^2	8.07	29.78	<0.0001	817.84	29.81	<0.0001
B^2	0.21	0.77	0.39	92.76	3.38	0.08
C^2	0.01	0.04	0.85	178.42	6.50	0.02
Residual	4.88			493.77		
Cor total	229.25			36 783.71		

SME20 blends are found to be approximately 6%, 11%, 9% and 12%, respectively, compared to the diesel fuel.

4.2.2 BTE model. The ANOVA result for the BTE is listed in Table 7. With an *F*-value of 91.94 and a corresponding *p*-value of less than 0.0001, the analysis of variance (ANOVA) for the quadratic response surface model of brake thermal efficiency (BTE) reveals that the model is really highly significant. This suggests that a good amount of the response's fluctuation is explained by the model. While the adjusted *R*-squared (0.9681) and predicted *R*-squared (0.9522) values are in great agreement, verifying the predictive power of the model, the low residual sum of squares (4.88) and high *R*-squared value (0.9787) point to a solid match. While other components like *A*, *B*, and *C* (brake power) are not significant, the important term A^2 (speed squared) with a *p*-value of less than 0.0001 reveals a quadratic relationship with BTE. Confirming that the model is sufficient for design space exploration, the accuracy of precision—measuring the signal-to-noise ratio—is 30.362, much higher than the intended threshold of 4.

$$\mathrm{BTE} = 478.5 + 0.03\mathrm{speed} - 23.99\mathrm{LCV} - 0.155\mathrm{BP} - 0.00039\mathrm{speed} \times \mathrm{LCV} + 0.0004\mathrm{speed} \times \mathrm{BP} - 0.0145\mathrm{LCV} \times \mathrm{BP} - 0.0000038\mathrm{speed}^2 + 0.31\mathrm{LCV}^2 - 0.004\mathrm{BP}^2$$

The variation in the BTE with a change in engine speed is presented in Fig. 4. The BTEs of biodiesel fuels are observed to be lower than those of neat diesel. The possible reason is the higher viscosity and density and lower CV of biodiesel fuel compared to neat diesel. The same trend for the BTE was explained by Nantha *et al.*,[59] who conducted an experiment in a 4.4 kW, constant-speed, four-stroke DI diesel engine running on biodiesel produced from waste cooking oil in the ratio ranging from B20 to B100. With an increase in engine speed, the BTE value increases but up to 3000 rpm; after that, it shows



(a) BTE vs. speed

(b) BTE vs. LCV

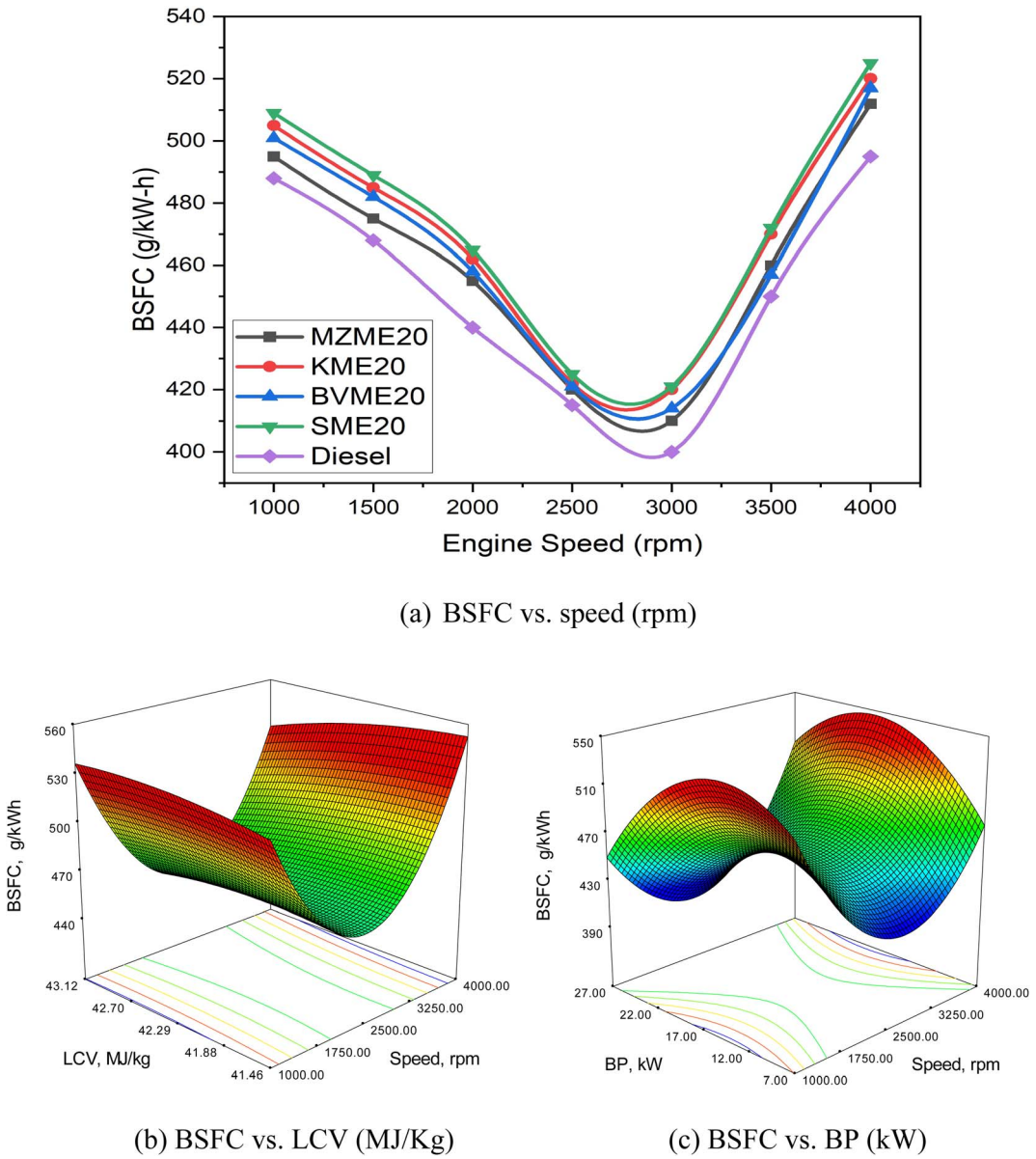
(c) BTE vs. BP (kW)

Fig. 4 Brake thermal efficiency variation with engine speed: (a) BTE *vs.* speed, (b) BTE *vs.* LCV and (c) BTE *vs.* BP (kW).

a decreasing trend. At 3000 rpm, the BTEs are reported to be 23.5%, 23%, 22.8%, 22.6% and 22.3% for diesel, MZME20, BVME20, KME20 and SME20, respectively. Among the different biodiesel feedstocks, MZME20 shows a superior BTE. This is attributed to the higher volatility and lower viscosity of MZME, which leads to better atomization and results in better mixing. The blends MZME20, BVME20, KME20 and SME20 present the average reductions in the BTE by around 2%, 6%, 4% and 8%, respectively, compared to the diesel fuel.

4.2.3 BSFC model. The ANOVA outcomes for BSFC are listed in Table 7. With an F-value of 146.99 and a p-value of less than 0.0001, the ANOVA result for the quadratic response surface model for BSFC shows that the model is significant. This implies that the reported model F-value may be the result of random noise with a really low probability (0.01%). Significant contributions among the model terms are the interaction term BC (p-value = 0.0105) and the quadratic terms A^2 (p-value = 0.0001) and C^2 (p-value = 0.0201). Other terms, such as speed (A) and brake power (C), show larger p-values, showing their lower relevance in the model. The high R-squared result (0.9866) indicates that the model explains 98.66% of the variability in the response; the adjusted R-squared value (0.9799) fits very nicely with the expected R-squared value (0.9711), hence suggesting great predictive power. The suitable precision ratio of 36.333 further supports the resilience of the model for traversing the design space.

$$\begin{aligned}\text{BSFC} = {} & -10017.86 + 0.05\text{speed} + 522.7\text{LCV} - 62.48\text{BP} \\ & - 0.006\text{speed} \times \text{LCV} + 0.0016\text{speed} \times \text{BP} + 1.8\text{LCV} \\ & \times \text{BP} + 0.000038\text{speed}^2 - 6.42\text{LCV}^2 - 0.54\text{BP}^2\end{aligned}$$



(a) BSFC vs. speed (rpm)

(b) BSFC vs. LCV (MJ/Kg)

(c) BSFC vs. BP (kW)

Fig. 5 Brake specific fuel consumption variation with engine speed: (a) BSFC *vs.* speed (rpm), (b) BSFC *vs.* LCV (MJ kg^{-1}) and (c) BSFC *vs.* BP (kW).

BSFC variations of all test fuels at different speeds are compared and shown in Fig. 5. BSFC is observed to decrease with an increase in engine speed up to 3000 rpm, and beyond this speed, BSFC again increases for all the test fuels. The diesel fuel shows a lower BSFC compared to other biodiesel fuels. The minimum value of BSFC is obtained at 3000 rpm for MZME20, which is 410 g kW^{-1} h^{-1}, followed by BVME20, KME20 and SME20, which are obtained around 414, 419 and 421 g kW^{-1} h^{-1}, respectively. The density of biodiesel is higher than that of the diesel fuel, requiring a larger mass of fuel injection for the same volume. This results in an increase in fuel consumption compared to the diesel fuel. The same behaviour has also been mentioned by other researchers[60,61] who have conducted experiments in a single-cylinder, naturally aspirated, four-stroke, water-cooled, direct-injection diesel engine fuelled with B0–B100 blends of soybean biodiesel. Fig. 6 shows the average increase in BSFC for MZME20, BVME20, KME20 and SME20 blends, which were approximately 2%, 4%, 3% and 5% than those of the diesel fuel, respectively.

4.2.4 CO emission model. The ANOVA results for emissions like CO, HC and NO_x are listed in Table 8. Having an *F*-value of 64.25 and a *p*-value of less than 0.0001, the study of variance (ANOVA) for the quadratic model projecting carbon monoxide (CO) emissions reveals that the model is quite significant. This proves the dependability of the model because it shows a quite low (0.01%) likelihood of attaining such a high *F*-value due to noise. With a *p*-value of 0.0360, the variable

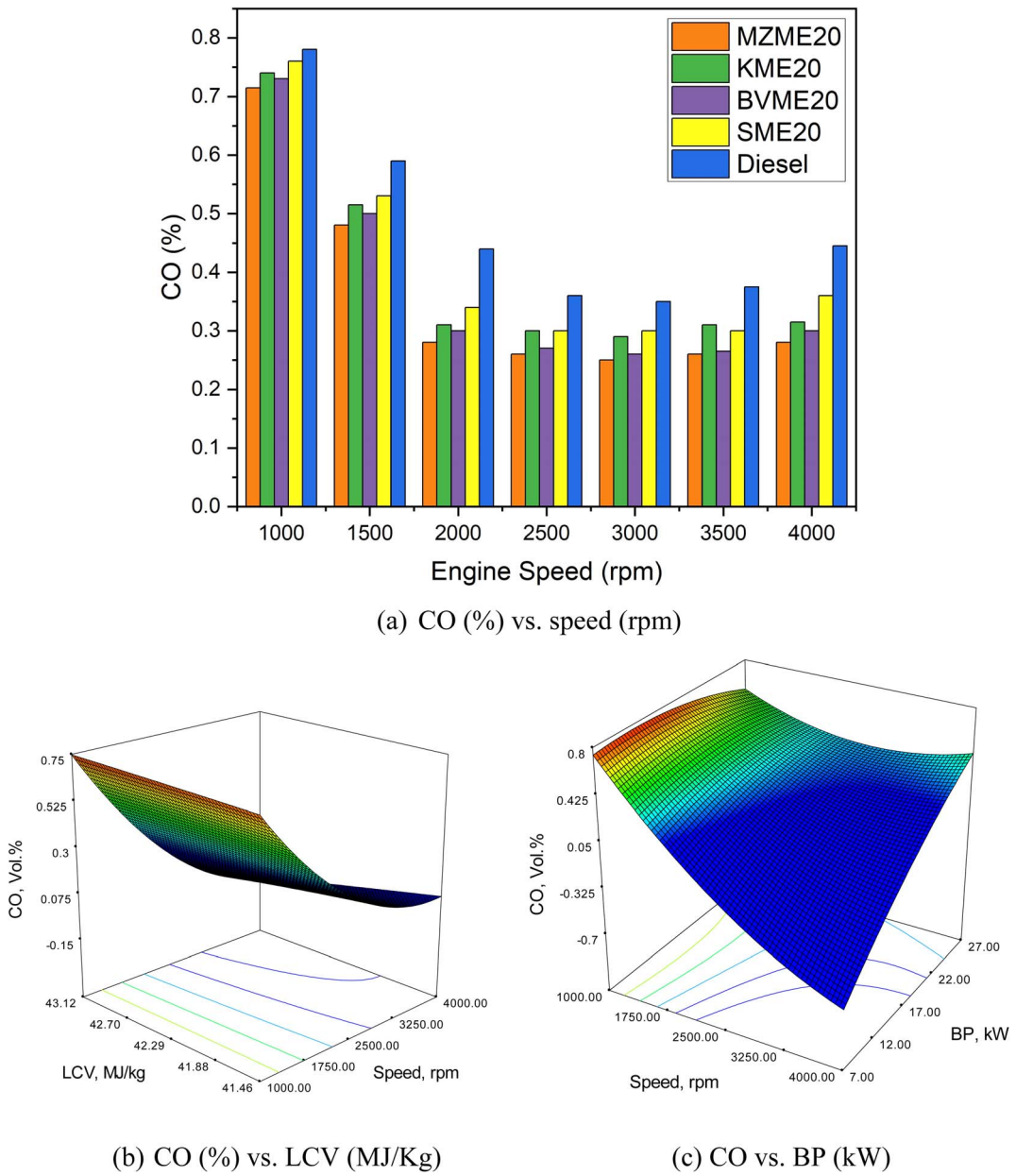


Fig. 6 CO (%) emission variation with engine speed: (a) CO (%) *vs.* speed (rpm), (b) CO (%) *vs.* LCV (MJ kg^{-1}) and (c) CO *vs.* BP (kW).

Table 8 ANOVA results for engine emissions

	CO			HC			NO_x		
Source	SS	*F*-Value	*p*-Value	SS	*F*-Value	*p*-Value	SS	*F*-Value	*p*-Value
Model	0.666	64.250	<0.0001	274.683	240.798	<0.0001	47 063.616	357.814	<0.0001
A-speed	0.004	3.509	0.077	0.289	2.280	0.148	93.223	6.379	0.021
B-LCV	0.006	5.136	0.036	3.969	31.312	<0.0001	614.797	42.067	<0.0001
C-BP	0.001	1.130	0.302	0.032	0.255	0.620	1.173	0.080	0.780
AB	0.003	2.983	0.101	0.164	1.296	0.270	33.082	2.264	0.150
AC	0.002	1.825	0.193	0.001	0.011	0.917	9.292	0.636	0.436
BC	0.002	1.604	0.221	0.026	0.203	0.657	30.222	2.068	0.168
A^2	0.004	3.351	0.084	0.033	0.263	0.615	125.092	8.559	0.009
B^2	0.000	0.007	0.933	1.099	8.672	0.009	47.667	3.262	0.088
C^2	0.001	0.445	0.513	0.025	0.198	0.662	2.831	0.194	0.665
Residual	0.021			2.281			263.062		
Cor total	0.687			276.964			47 326.679		

indicating the lower calorific value (LCV) among the model terms is the only one having a significant impact on CO emissions. Not statistically significant terms include the speed ($p = 0.0774$), braking power (BP, $p = 0.3019$), interactions such as *AB*, *AC*, and *BC*, and quadratic terms (A^2, B^2, and C^2). The adjusted *R*-squared value (0.9547) is quite similar to the highest observed *R*-squared value of 0.9698, suggesting that about 97% of the variance in CO emissions can be explained by the independent variables; the model accuracy after accounting for the number of predictors. Furthermore, displaying a significant predictive power is the projected *R*-squared value of 0.9159. The appropriate precision ratio of 24.062 exceeds the intended threshold of 4, thereby verifying that the model has a strong signal-to-noise ratio and is fit for design optimization. The coefficient of variation (CV) of 8.72% shows moderate variability.

$$\begin{aligned}\mathrm{CO} = {} & -0.55 + 0.00089\mathrm{speed} + 0.132\mathrm{LCV} - 0.22\mathrm{BP} \\ & - 0.000046\mathrm{speed} \times \mathrm{LCV} + 0.000022\mathrm{speed} \times \mathrm{BP} \\ & + 0.005\mathrm{LCV} \times \mathrm{BP} + 8.2 \times 10^{-8}\mathrm{speed}^2 \\ & - 0.0019\mathrm{LCV}^2 - 0.00092\mathrm{BP}^2\end{aligned}$$

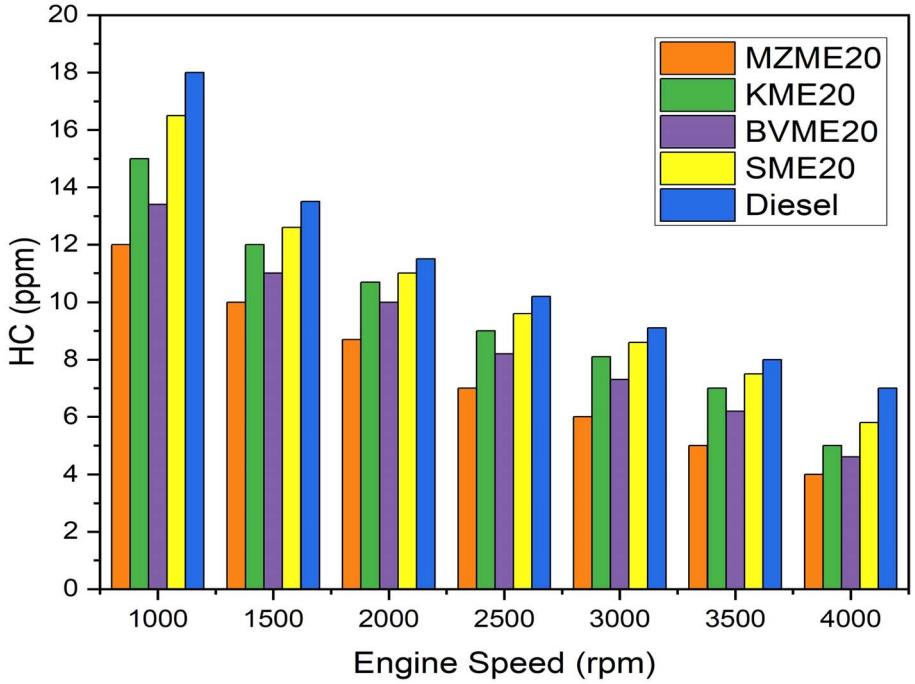
Incomplete combustion is the primary cause of CO emissions from engines. Fig. 6 shows the CO emission pattern with the complete range of engine speed. The CO emission is lower for biodiesel blends than for the diesel fuel. The higher cetane number, low carbon content and higher oxygen content of biodiesel fuel compared to diesel are the possible reasons for its lower CO emissions, as stated by Xue *et al.*[61] Hirkude and Padalkar[62] reported that the higher oxygen content of biodiesel allows a large number of carbon molecules to participate in burning, resulting in better combustion. The CO emission can also be reduced by increasing the injection timing. Lower CO emissions are obtained at high air-fuel ratios and high combustion temperatures.[50–56] At a speed of 3000 rpm, the CO emissions for MZME20, BVME20, KME20, SME20 and diesel are observed to be 0.25%, 0.26%, 0.29%, 0.29% and 0.38%, respectively. The average reductions in the emissions of CO for MZME20, BVME20, KME20 and SME20 blends are found to be approximately 3%, 2%, 3% and 2%, respectively, compared to the diesel fuel.

4.2.5 HC emission model. In the case of the HC emission model having an *F*-value of 240.80, the ANOVA findings for the HC emission model expose a very significant quadratic model that indicates the great capacity of the model to explain the data variability (Table 8). With a *p*-value of <0.0001, the related probability that this significant *F*-value might arise from random noise has less than a 0.01% chance. With *p*-values of 0.0001 and 0.0087, respectively, the LCV (*B*) and its quadratic term (B^2) are major players in the model because fluctuations in the low calorific value (LCV) clearly influence HC emissions. Little terms like speed (*A*) and BP (*C*) imply their little influence on the HC emission forecast. While the predicted *R*-squared value of 0.9801 further verifies its robustness, the *R*-squared value of 0.9918 and adjusted *R*-squared of 0.9876 show great accuracy and fit to the data. The model, appropriate for design optimization, shows a robust signal-to-noise ratio with a high Adeq precision of 56.522.

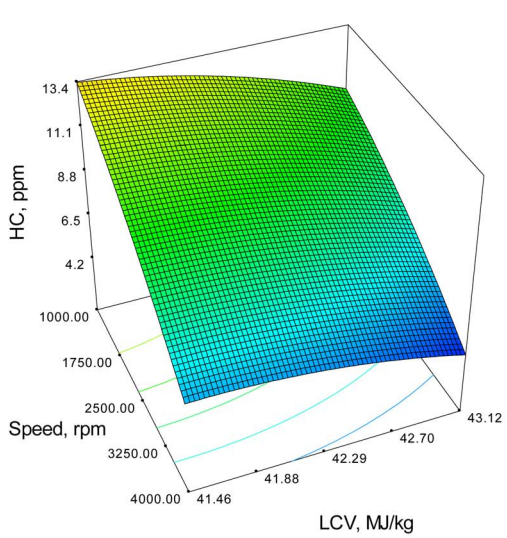
$$\begin{aligned}\mathrm{HC} = {} & -1148.79 - 0.015\mathrm{speed} + 57.13\mathrm{LCV} + 0.44\mathrm{BP} \\ & + 0.0003\mathrm{speed} \times \mathrm{LCV} + 0.00002\mathrm{speed} \times \mathrm{BP} \\ & - 0.019\mathrm{LCV} \times \mathrm{BP} - 2.41289 \times 10^{-7}\mathrm{speed}^2 \\ & - 0.7\mathrm{LCV}^2 + 0.006\mathrm{BP}^2\end{aligned}$$

Fig. 7 indicates the variation in HC emissions at different engine speeds for all test fuels. All the feedstocks taken under study show less HC emissions than neat diesel. This is due to the high oxygen content and higher cetane number of biodiesel fuel. On average, the percentage reduction in the HC emissions of MZME20, BVME20, KME20 and SME20 is found to be around 31%, 21%, 13% and 8%, respectively, compared to diesel. A lower HC emission is found for MZME20 compared to other feedstocks, which is due to the comparatively higher cetane number of MZME20. The similar behaviour was also mentioned in.[63]

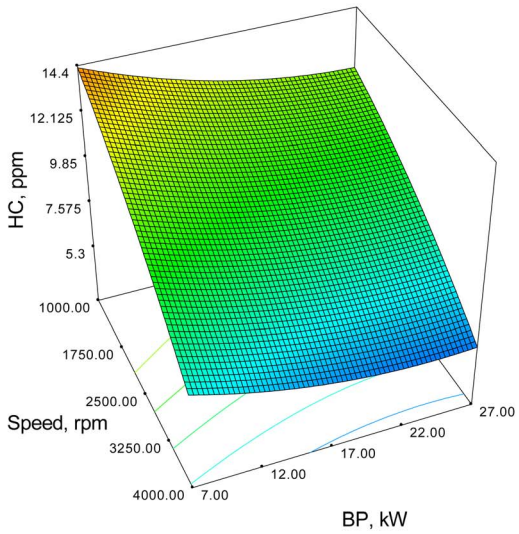
4.2.6 Nitrogen oxide emissions. With an *F*-value of 357.81 and a *p*-value less than 0.0001, the ANOVA table for the NO_x emission quadratic response surface model shows that the model is highly significant—that is, there is less than 0.01% probability that such a big *F*-value could arise from random



(a) HC (ppm) vs. speed (rpm)



(b) HC (ppm) vs. LCV (MJ/Kg)



(c) HC (ppm) vs. BP (kW)

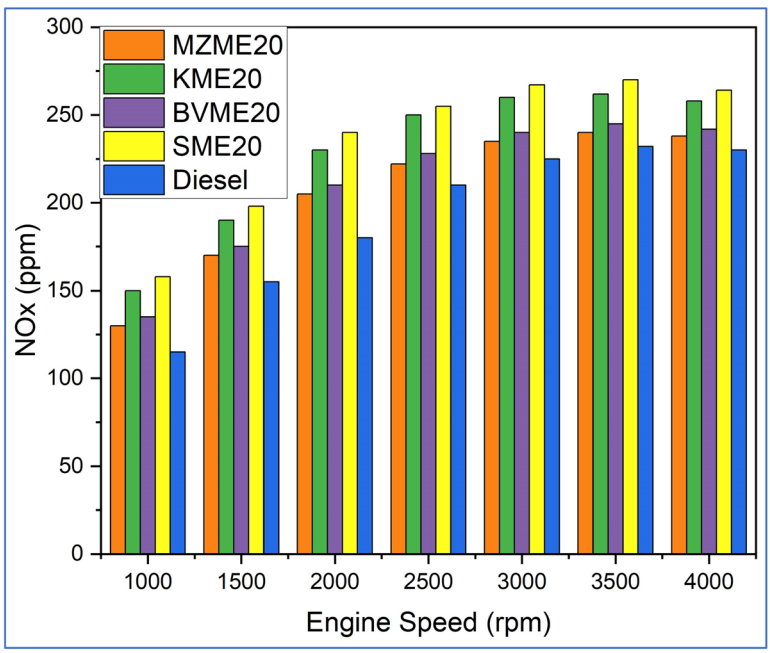
Fig. 7 HC emissions: (a) HC (ppm) *vs.* speed (rpm), (b) HC (ppm) *vs.* LCV (MJ kg^{-1}) and (c) HC (ppm) *vs.* BP (kW).

noise (Table 8). The statistically significant (p-values < 0.05) key elements influencing the NO_x emissions are engine speed (A), reduced calorific value (B), and quadratic term of the speed (A^2). By contrast, with p-values considerably higher than 0.05, brake power (C) and most interaction terms (AB, AC, and BC) are not significant. With an R-squared value of 0.9944 and an adjusted R-squared value of 0.9917, the model shows great predictive ability (Pred R-squared = 0.9879). Further indicating the model's ability to efficiently traverse the design space and generate consistent forecasts are the low coefficient of variance (1.73%) and high adequate precision (63.002).

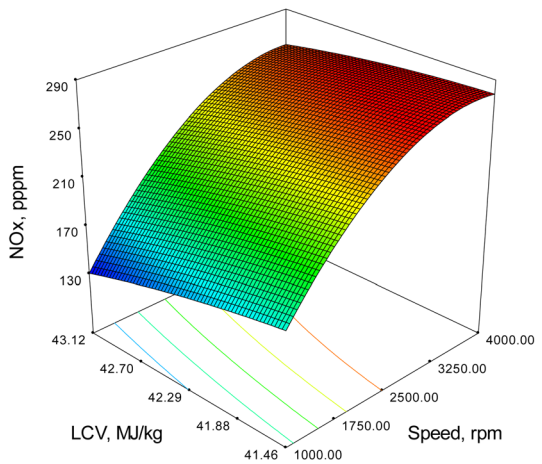
$$NO_x = -7419.9 - 0.052\text{speed} + 370.56\text{LCV} + 28.53\text{BP} + 0.0045\text{speed} \times \text{LCV} - 0.0012\text{speed} \times \text{BP} - 0.66\text{LCV} \times \text{BP} - 0.000015\text{speed}^2 - 4.6\text{LCV}^2 + 0.07\text{BP}^2$$

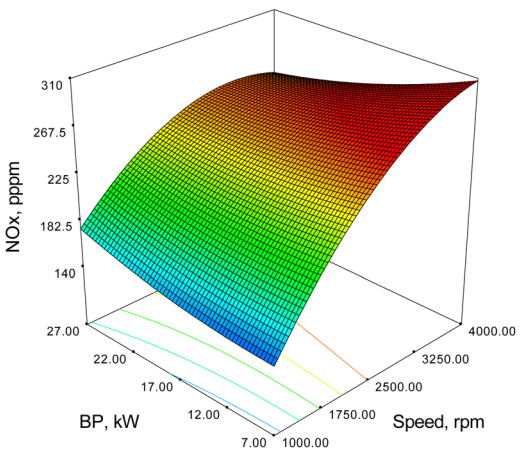
Palash *et al.*[64] reported that NO_x formation is strongly reliant on the in-cylinder temperature, oxygen concentration and reaction residence time. Fig. 8 presents the variation in NO_x emissions for all test fuels with a change in engine speed. NO_x emissions are reported to be higher for all biodiesel feedstocks compared to neat diesel across the entire speed range. This is due to the high oxygen content of biodiesel fuel. Also, the cetane number of biodiesel is higher, which is accountable for the decrease in the ignition delay and, in turn, the improvement in the combustion that results in higher NO_x emissions, as explained by Xue *et al.*[61] NO_x emissions for MZME20, BVME20, KME20 and SME20 are found to higher on an average by around 6%, 9%, 18% and 20%, respectively, than the diesel fuel. NO_x emissions also depend on the adiabatic flame temperature, which, in turn, is a function of the unsaturated fatty acid content in biodiesel fuel, as is well supported by other literature studies.[65,66]

The observed performance and emission variations among biodiesel blends correlate strongly with their fatty acid compositional profiles (Table 1). MZME20 demonstrates the highest BTE among biodiesel blends and optimal emission



(a) NOx vs. speed (rpm)





(b) NO_x vs. LCV (MJ/Kg) (c) NO_x vs. BP (kW)

Fig. 8 NO_x emission variation with engine speed: (a) NO_x *vs.* speed (rpm), (b) NO_x *vs.* LCV (MJ kg^{-1}) and (c) NO_xO *vs.* BP (kW).

characteristics (lowest CO, HC, and NO_x), which can be attributed to its oleic acid dominance (63.83%) and high unsaturation (83.54%), which enhances fuel atomization, extends ignition delay, and reduces peak combustion temperatures. SME20 shows the lowest BTE among biodiesel blends despite a high cetane number (57.4), potentially due to its higher viscosity affecting spray characteristics, though it exhibits higher NO_x emissions due to its high saturation (42.1%), promoting rapid, high-temperature combustion. The NO_x formation mechanisms follow the temperature-dependent Zeldovich thermal pathway, where MZME20's extended ignition delay creates lower peak flame temperatures, suppressing NO formation, while SME20's shorter ignition delay results in rapid heat release with locally elevated temperatures, promoting thermal NO_x generation. The higher BSFC of all biodiesel blends reflects lower calorific values than those of diesel, while reduced CO and HC emissions across all blends stem from their inherent oxygen contents (~11% by mass), facilitating complete oxidation.[67,68] These structure–property relationships demonstrate that the compositional balance, particularly oleic acid content and degree of saturation, significantly influences the combustion efficiency and emission formation mechanisms.[69]

4.2.7 Desirability-based optimization of the engine performance. Presenting the desirability-based optimization of engine performance, Table 9 emphasizes the ideal parameter choices to attain a balanced performance at several engine outputs.[70] While the lower calorific value (LCV) is maximized at 43.12 MJ kg^{-1}, the braking power (BP) is near its maximum limit at 26.42 kW, and the speed is optimal at 2852 rpm. Although brake-specific fuel consumption (BSFC) is the lowest at 413.67 g kW^{-1} h^{-1}, somewhat below the lower limit, the brake thermal efficiency (BTE) reaches its highest feasible value of 23%. Emissions, including carbon monoxide (CO), hydrocarbons (HC), and nitrogen oxides (NO_x), are reduced to 0.275 g kW^{-1} h^{-1}, 6.5 g kW^{-1} h^{-1}, and 230 g kW^{-1} h^{-1}, respectively, thereby suggesting a development in emission control. The optimized settings and results are presented in Table 9.

Fig. 9 gives the desirability bar plots obtained using the powerful desirability-based optimization approach. The desirability bar plots in Fig. 9 illustrate how well each response variable (NO_x, HC, CO, BTE, and BSFC) meets the desired optimization goals. In a desirability-based optimization approach, each output is assigned a desirability score between 0 (completely undesirable) and 1 (fully desirable), and a combined desirability score reflects the overall effectiveness of the optimization.[71]

Table 9 Optimized settings and results

Name	Goal	Lower limit	Upper limit	Optimized results
Speed	Is in range	1000	4000	2852
LCV	Is in range	41.46	43.12	43.12
BP	Is in range	7	27	26.42
BTE	Maximize	14	23	23
BSFC	Minimize	414	526	413.67
CO	Minimize	0.2	0.8	0.275
HC	Minimize	4	16	6.5
NO_x	Minimize	130	271	230
Desirability				0.835

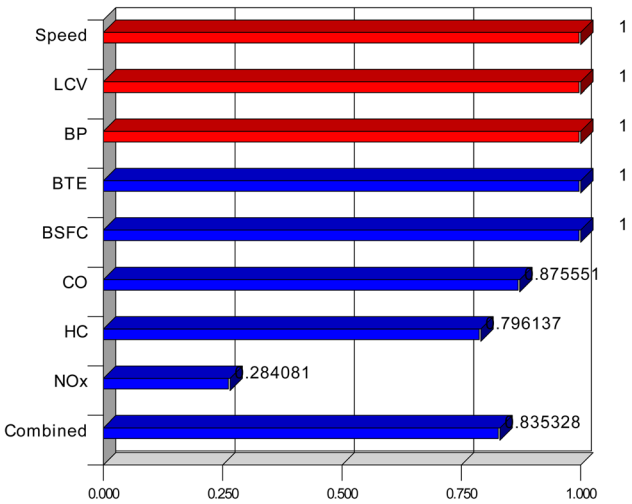


Fig. 9 Desirability bar plots.

HC – 0.79: a fairly high score, indicating that unburned hydrocarbon emissions are significantly reduced and close to the target. CO – 0.875: a high desirability score, showing that carbon monoxide emissions are effectively minimized. Brake thermal efficiency (BTE) – 1: the maximum score indicates the engine achieves the highest possible thermal efficiency under the given conditions. Brake specific fuel consumption (BSFC) of 1 is also at the optimal level, meaning the fuel efficiency is maximized.

The combined desirability score of 0.835 indicates a high trade-off between performance and emission minimization, thereby indicating that the optimization attains a balanced and favourable engine design within the given limitations.

4.3 ML modelling

The dataset comprises experimental measurements from a multi-cylinder diesel engine using 20% biodiesel blends at eight engine speeds for four feedstocks. For model evaluation, we employed fivefold cross-validation to ensure robustness and minimize overfitting, with data folds randomly and evenly divided. Pre-processing included normalization and correlation-based feature selection. Detailed train-test splits were not used separately, as cross-validation sufficed.

4.3.1 Correlation analysis of data. Emphasizing the speed (rpm), load (kW), thermal efficiency (BTE), fuel consumption (BSFC), and emissions (CO, HC, and NO_x), the correlation matrix (Table 10) and correlation heat map (Fig. 10) reveal links among engine characteristics. Given a strong positive connection between the speed and load—0.89—the load usually increases as engine speed increases. While the speed exhibits a substantial negative association with hydrocarbon (HC) emissions (−0.93), showing that greater speeds reduce unburnt hydrocarbons but increase nitrogen oxides, the speed is also favorably connected with NO_x (0.86). The load similarly shows strong positive associations with the BTE (0.93) and NO_x (0.91)

Table 10 Correlation matrix

	Speed, rpm	LCV, MJ kg^{-1}	Load, kW	BTE, %	BSFC, g kW^{-1} h^{-1}	CO, vol%	HC, ppm	NO_x, ppm
Speed, rpm	1	0	0.89	0.75	−0.07	−0.74	−0.93	0.86
LCV, MJ kg^{-1}	0	1	0.1	0.14	−0.09	−0.09	−0.31	−0.3
Load, kW	0.89	0.1	1	0.93	−0.49	−0.92	−0.91	0.91
BTE, %	0.75	0.14	0.93	1	−0.68	−0.84	−0.79	0.81
BSFC, g kW^{-1} h^{-1}	−0.07	−0.09	−0.49	−0.68	1	0.55	0.18	−0.39
CO, vol%	−0.74	−0.09	−0.92	−0.84	0.55	1	0.8	−0.87
HC, ppm	−0.93	−0.31	−0.91	−0.79	0.18	0.8	1	−0.75
No_x, ppm	0.86	−0.3	0.91	0.81	−0.39	−0.87	−0.75	1

but negatively correlates with CO (−0.92), hence implying improved combustion at greater loads. Fascinatingly, the brake specific fuel consumption (BSFC) is inversely linked to the BTE (−0.68), thereby demonstrating better fuel economy at higher thermal efficiency levels. The lower calorific value (LCV) of the fuel shows poor associations with other variables, thereby suggesting little direct impact on performance or emissions in this dataset.[72]

4.3.2 Model development and evaluation

4.3.2.1 BTE model development using ML algorithms. The models' performance measures on the BTE data show different traits in generalizing the capacity and prediction accuracy (Fig. 11). With a training mean squared error (MSE) of 0.84 and a testing MSE of 0.67, the linear regression model shows a quite strong fit with a training *R*-squared value of 0.8834 and a higher testing *R*-squared of 0.9309 (Table 11). This shows good predictive ability because the model explains almost 93% of the variance in the test set. By contrast, the decision tree model attains an amazing training *R*-squared of 1.0000 and a flawless training MSE of 0.00, thereby indicating robust learning. The decision tree for BTE was optimized with a depth of 12, a split threshold of 8, and a leaf size of 3, as shown in Table 12. This

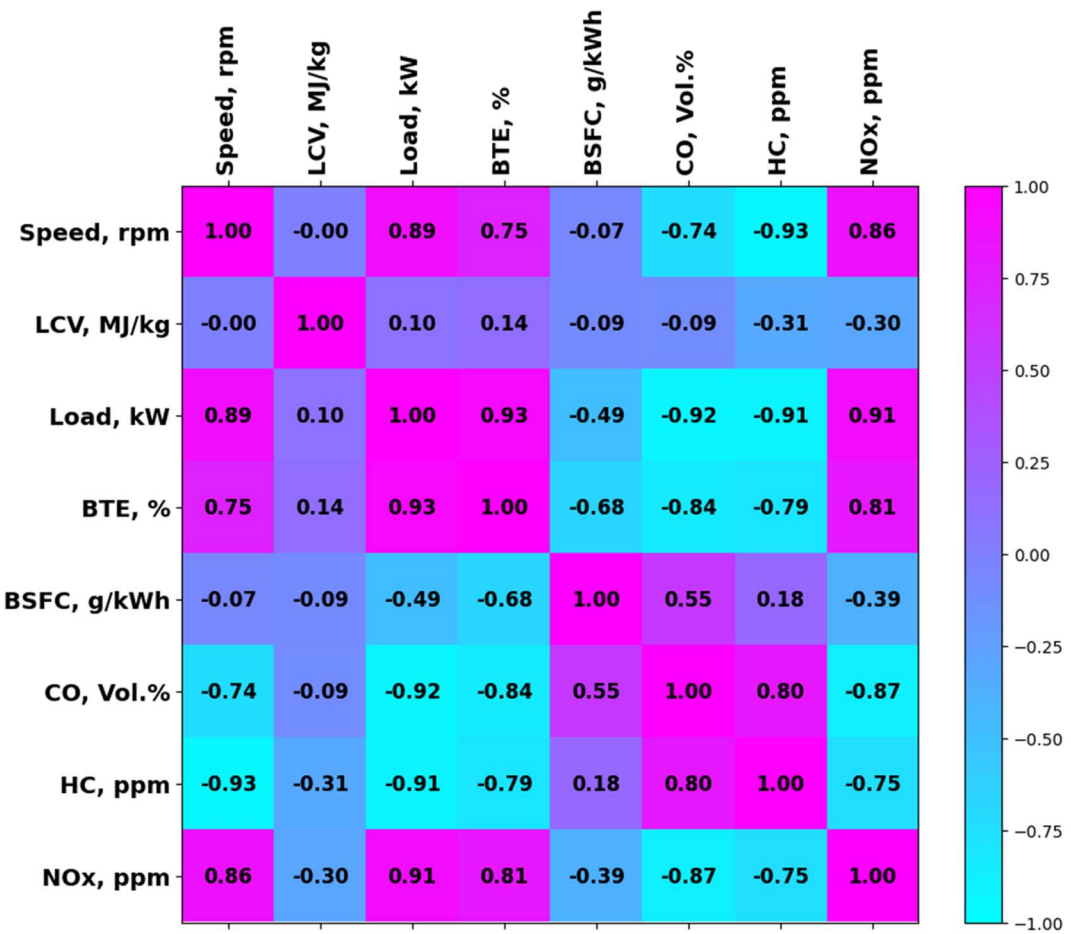


Fig. 10 Correlation heat map.

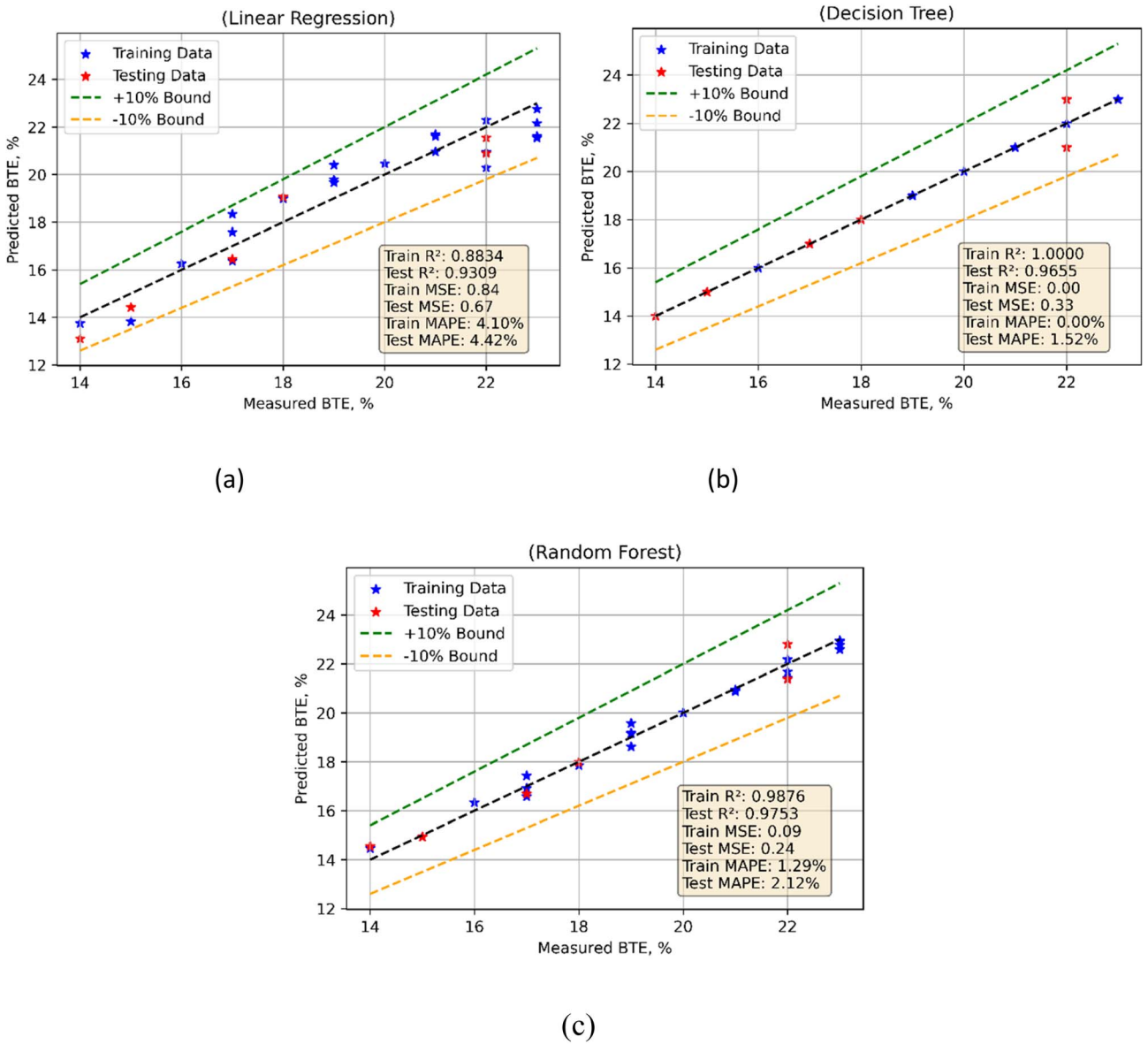


Fig. 11 BTE measured *vs.* forecasted values in the cases of (a) LR, (b) DT and (c) RF models.

setup shows that a reasonable amount of the tree depth is enough to capture the nonlinearity linked to efficiency without too much branching. The ensemble for Random Forest chose 120 trees with a maximum depth of 25 and "sqrt" feature sampling. Random Forest is more sophisticated than other models, which shows that it can handle more delicate interactions between various input variables. Its testing MSE of 0.33 and testing *R*-squared of 0.9655, however, show a notable decline in performance, suggesting over-fitting. With *R*-squared values of 0.9876 for training and 0.9753 for testing, Random Forest shows a training MSE of 0.09 and a testing MSE of 0.24, thereby indicating its strong generalizing powers. Furthermore, the Random Forest model shows lower mean absolute percentage errors (MAPEs) than the other models, thereby verifying its superiority in predictive modelling in terms of the generalization and balance of accuracy.

Table 11 Statistical evaluation of models

		Train	Test	Train	Test	Train	Test
ML	Model	MSE	MSE	R^2	R^2	MAPE, %	MAPE, %
LR	BTE	0.84	0.67	0.8834	0.9309	4.10	4.42
DT		0.00	0.33	1	0.9655	0.00	1.52
RF		0.09	0.24	0.9876	0.9753	1.29	2.12
LR	BSFC	103.61	221.07	0.9279	0.7338	1.92	2.55
DT		0.00	42.17	1	0.9492	0.00	1.03
RF		4.67	46.62	0.9967	0.9439	0.39	1.19
LR	CO	0.00	0.01	0.8654	0.8523	11.97	13.68
DT		0.00	0.00	1	0.9597	0.00	2.08
RF		0.00	0.00	0.9879	0.9328	2.68	4.32
LR	HC	0.19	0.38	0.9788	0.9583	4.47	5.36
DT		0.00	1.17	1	0.8735	0.00	8.11
RF		0.14	1.93	0.9844	0.7909	4.13	10.16
LR	NO_x	37.06	40.12	0.9721	0.9830	2.40	3.10
DT		0.00	59.83	1.0000	0.9746	0.00	3.73
RF		17.37	85.88	0.9869	0.9635	1.54	4.41

Table 12 Training hyperparameters, range, and optimum value

Model	Hyperparameter	Range	BTE	BSFC	CO	HC	NO_x
Decision tree	Max_depth	>3	12	10	15	13	11
	Min_samples_split	2 to 20	8	10	6	8	9
	Min_samples_leaf	1 to 10	3	4	2	4	3
Random forest	n_estimators	50 to 500+	120	100	150	130	140
	Max_features	'Auto', 'sqrt', numeric	'sqrt'	'Sqrt'	'Auto'	'Sqrt'	'Auto'
	Max_depth	None or 3 to 50	25	20	None	30	22
	Min_samples_split	2 to 20	6	8	5	4	7
	Min_samples_leaf	1 to 10	3	4	3	2	3

4.3.2.2 BSFC model development using ML algorithms. BSFC models' performance measures expose notable variations in model behavior and prediction accuracy, as depicted in Fig. 12. With a training mean squared error (MSE) of 103.61 and a testing mean squared error (MSE) of 221.07, the linear regression model shows a quite poor fit and generalizing capacity (Table 11). Although the testing *R*-squared value of 0.7338 indicates a significant decline in predictive performance on unseen data, most likely due to the model's sensitivity to noise and its linear assumptions, the training *R*-squared value

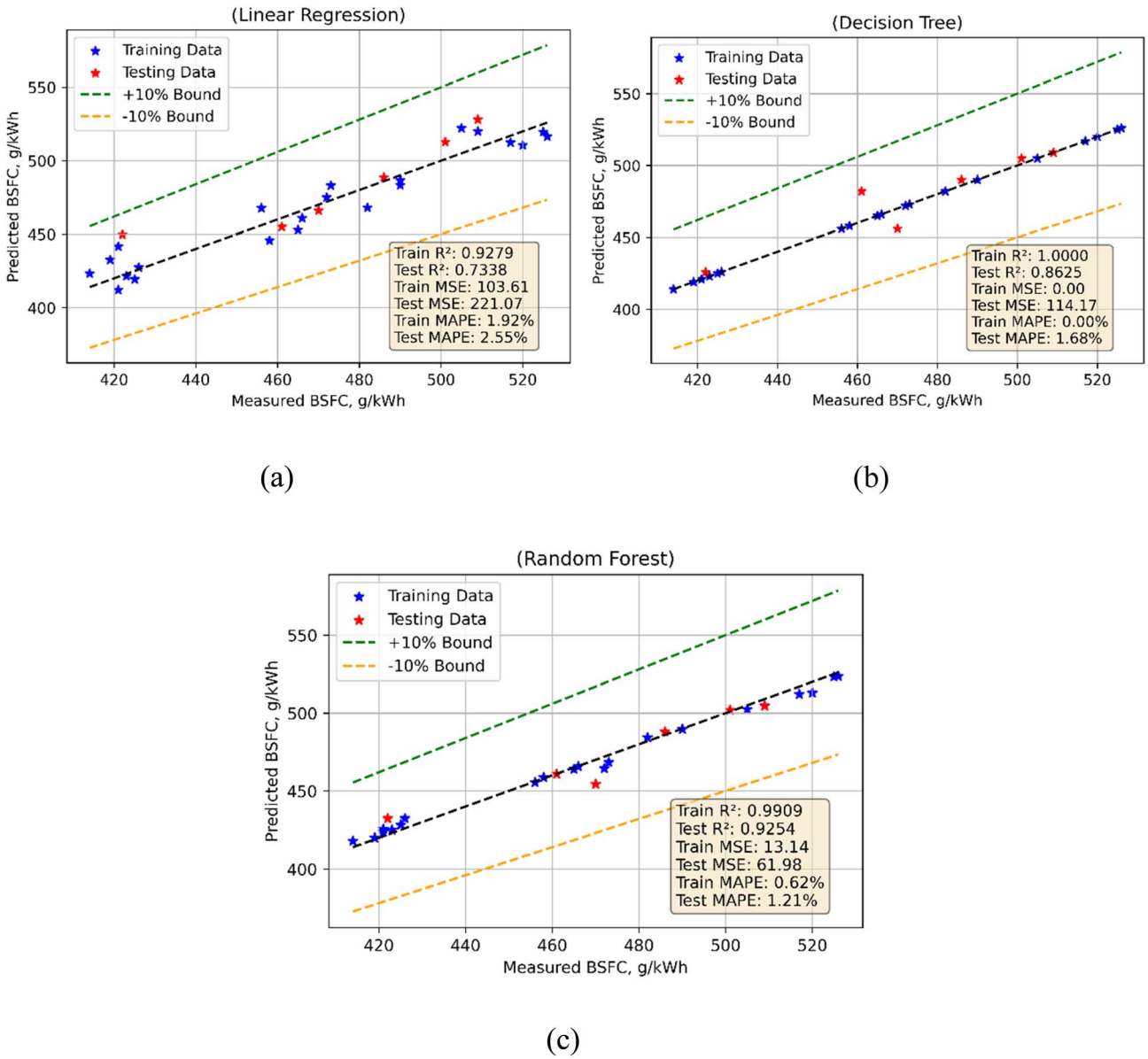


Fig. 12 BSFC measured *vs.* forecasted values in the cases of (a) LR, (b) DT and (c) RF models.

of 0.9279 indicates that approximately 93% of the variance in the training data is explained by the model. For the training hyperparameters' fine-tuning, the decision tree needed a depth of 10 for BSFC, and the splitting and leaf criteria were 10 and 4, respectively (Table 12). These numbers show that more control over node splitting is necessary to minimize overfitting while still getting fuel usage trends. Random Forest, by contrast, utilized 100 estimators with a maximum depth of 20 and "sqrt" features. The ensemble cut down on variance a lot, making it more stable for forecasting BSFC than the single tree. By contrast, the *R*-squared value of 1.0000 and a perfect training MSE of 0.00 indicate the total memorization of the training data in the decision tree model. Though still high, this results in a testing MSE of 42.17 and an *R*-squared value of 0.9492, which demonstrates overfitting as the model fails to broadly generalize effectively to fresh data. With a training MSE of 4.67 and a testing MSE of 46.62, the Random Forest model shows strong performance, with a training *R*-squared of 0.9967 and a testing *R*-squared of 0.9439. Among the three models, it is the most dependable because its lower training and testing mean absolute percentage errors (MAPEs) of 0.39 and 1.19, respectively, show its capacity to balance accuracy and generalisation.

4.3.2.3 CO emission model development using ML algorithms. The coefficients for the co-emission models show notable differences in predicting effectiveness among several modelling approaches (Fig. 13). With a training mean squared error (MSE) of 0.00 and a testing MSE of 0.01, the linear regression model shows that it performs practically exactly on the training data (Table 11). On the contrary, the training *R*-squared value of 0.8654 and testing *R*-squared value of 0.8523 point to a small

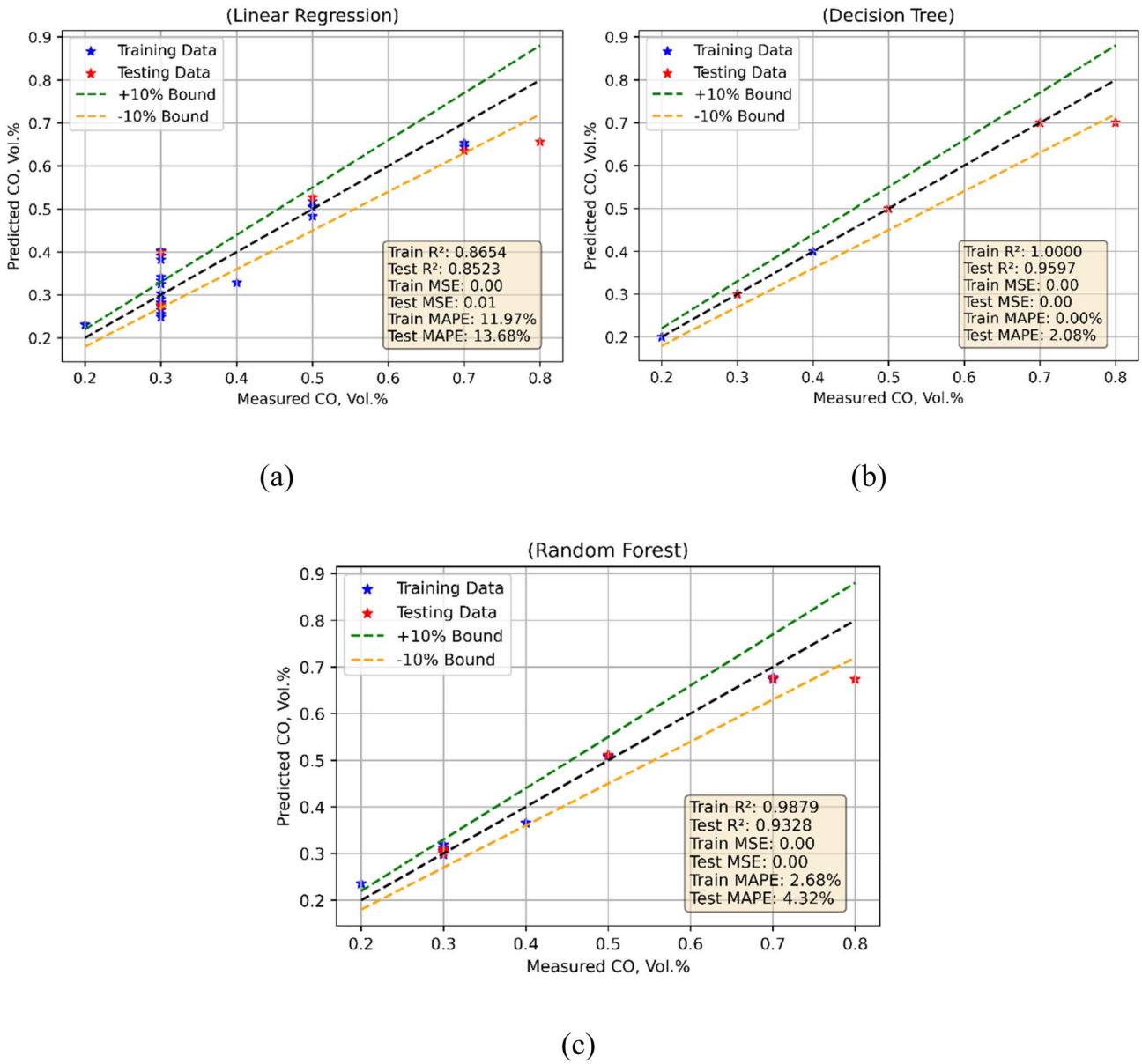


Fig. 13 CO emission measured *vs.* forecasted values in the cases of (a) LR, (b) DT and (c) RF models.

degree of explained variation; the model performs worse on unseen data. Particularly on new data, the somewhat high mean absolute percentage error (MAPE) values of 11.97 for training and 13.68 for testing highlight inconsistency in its forecasts. A perfect training *R*-squared of 1.0000 is obtained by the decision tree model with a perfect training MSE of 0.00 and a testing MSE of 0.00. This implies the total memorization of the training data; yet, the testing *R*-squared value of 0.9597 shows that, although not as successfully as the Random Forest model, the model still generalizes well. Last but not least, the Random Forest model shows a significant predictive capability while keeping low MAPE values of 2.68 and 4.32 for training and testing, respectively, with a high training *R*-squared value of 0.9879 and a testing *R*-squared value of 0.9328. Random Forest is a good choice for co-emission modelling because this emphasizes its strong generalization capacity. During training hyperparameters' fine-tuning, the CO emissions are more affected by deeper structures, as depicted in Table 12. The decision tree went 15 levels deep with a modest split value of 6 and a low leaf size of 2. This shows that more detailed splits are needed to show emission patterns. Random Forest was set up with 150 estimators and no limit on depth (none), and feature selection was set to "auto". This combination shows how important it is to capture complicated, high-dimensional interactions that affect CO production.

4.3.2.4 HC emission model development using ML algorithms. The three modeling approaches used show different degrees of predictive effectiveness according to the results for the

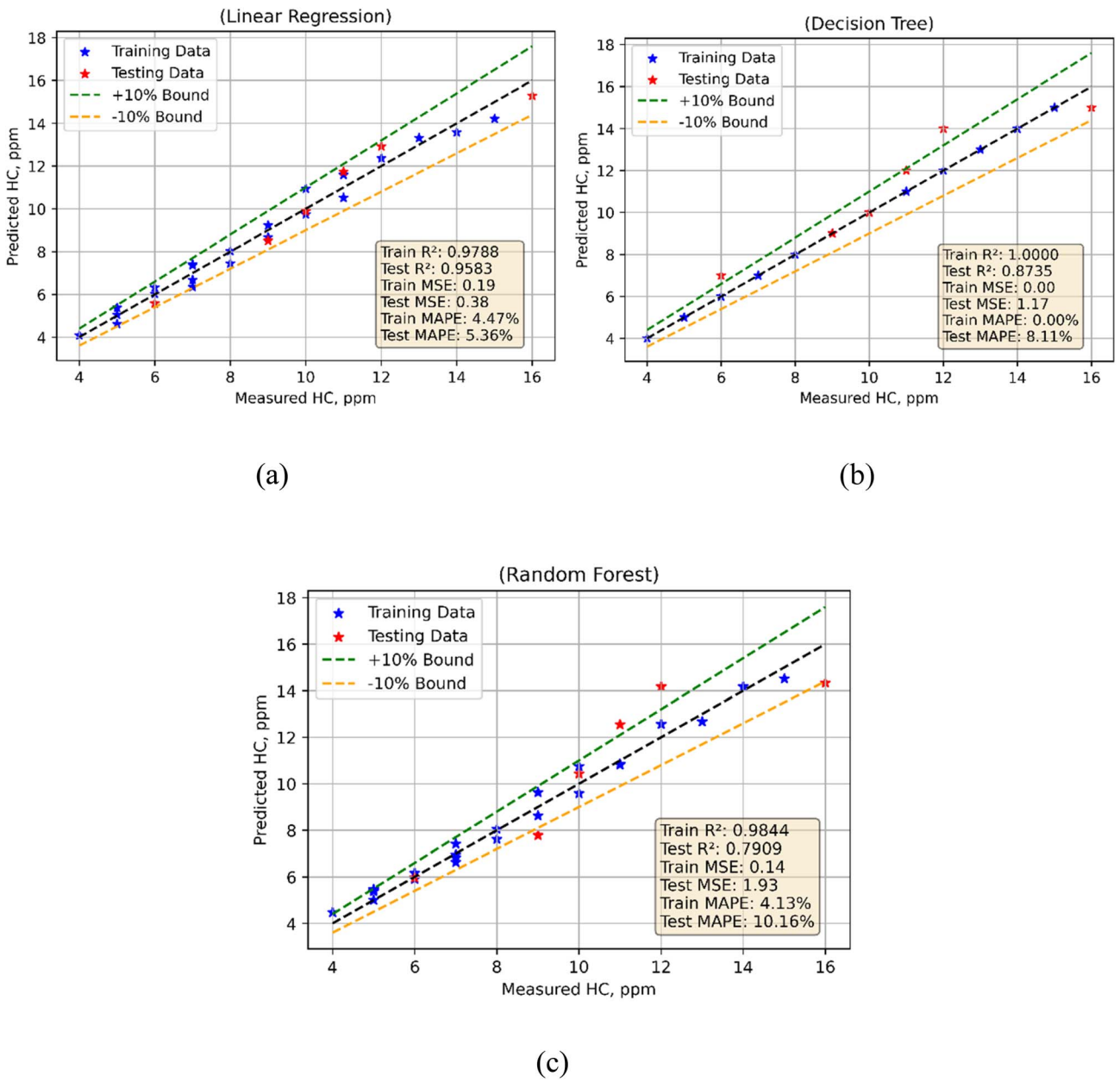


Fig. 14 HC emission measured *vs.* forecasted values in the cases of (a) LR, (b) DT and (c) RF models.

hydrocarbon (HC) emission model (Fig. 14). With a training mean squared error (MSE) of 0.19 and a testing mean squared error (MSE) of 0.38, the linear regression model shows a decent fit to the training set (Table 11). While the testing *R*-squared of 0.9583 indicates high generalization ability, albeit with some reduction in performance on untested data, the training *R*-squared value of 0.9788 indicates that approximately 98% of the variance in the training set is explained by the model. Further underlining its consistent predictive power are the mean absolute percentage error (MAPE) values of 4.47 for training and 5.36 for testing. The decision tree had a depth of 13 for HC emissions, with 8 splits and 4 leaves during fine-tuning of training hyperparameters. These numbers show that it is not too hard to show hydrocarbon release patterns. Random Forest used 130 estimators with a depth of 30, which made HC prediction more generic, as listed in Table 12. Using "sqrt" features helped keep the trees different while still getting useful predictions. On the contrary, the decision tree model exhibits a perfect training MSE of 0.00, hence suggesting the total memorization of the training set. Although it shows perfect predictive power with a training *R*-squared value of 1.0000, the testing MSE of 1.17 and a testing *R*-squared value of 0.8735 point to some overfitting as the model finds difficulty generalizing efficiently to fresh data.

With training and testing *R*-squared values of 0.9844 and 0.7909, respectively, the Random Forest model shows a training

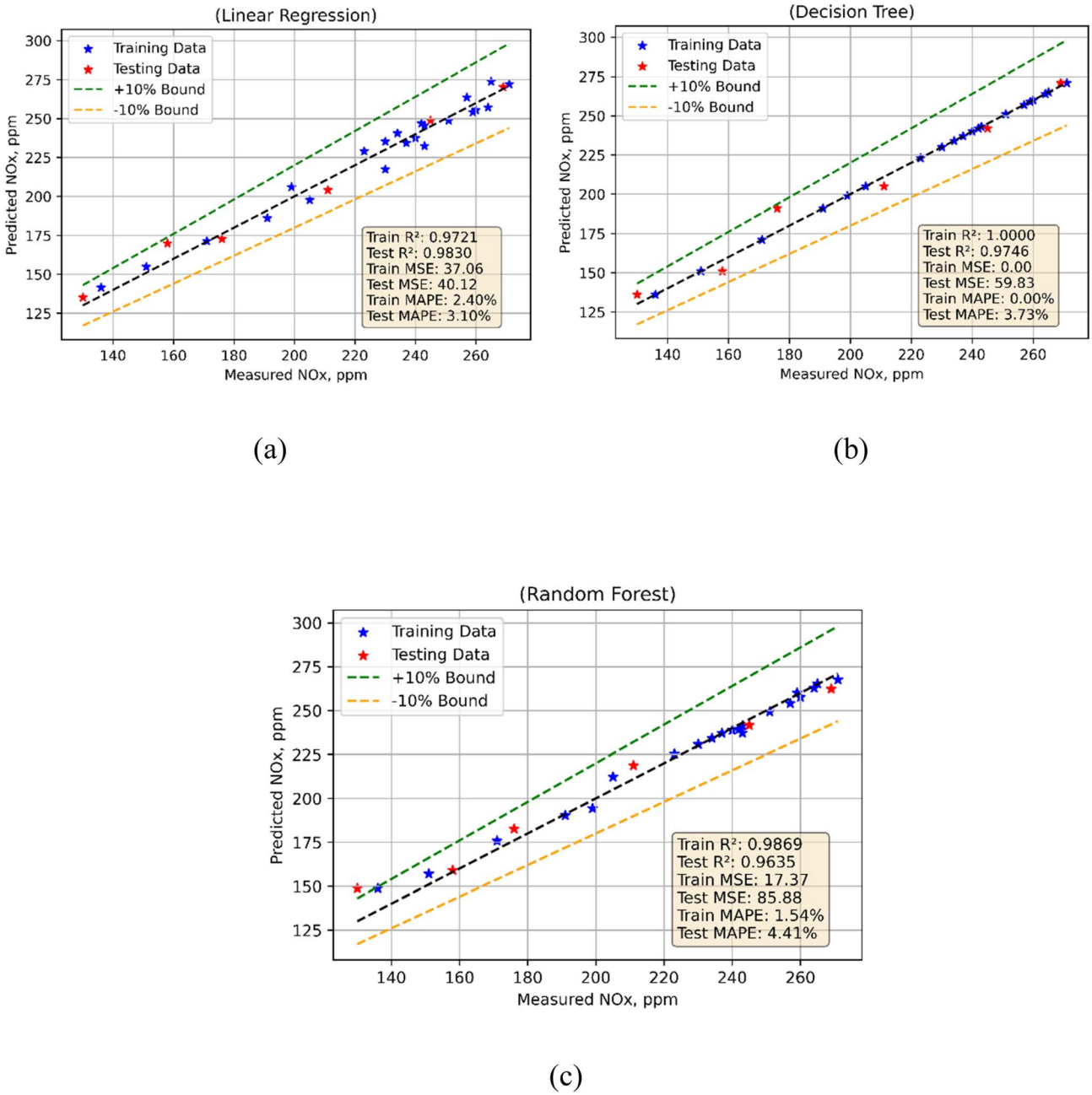


Fig. 15 NO_x emission measured *vs.* forecasted values in the cases of (a) LR, (b) DT and (c) RF models.

MSE of 0.14 and a testing MSE of 1.93. Although the decline in the testing *R*-squared value demonstrates a modest degree of overfitting, this shows strong performance. With MAPEs of 4.13 for training and 10.16 for testing, the model shows more unpredictability when forecasting on fresh data, even if it performs generally well.

4.3.2.5 NO_x emission model development using ML algorithms. Performance measures for the nitrogen oxide (NO_x) emission model expose different variations in predicted accuracy among the used techniques (Fig. 15). With a training mean squared error (MSE) of 37.06 and a testing mean squared error (MSE) of 40.12, the linear regression model exhibits. While the testing *R*-squared value of 0.9830 demonstrates high generalizing capacities on unknown data, the training *R*-squared value of 0.9721 indicates that around 97% of the variance in the training data is accounted for (Table 11). Furthermore, reflecting the model's dependability in generating predictions are the mean absolute percentage error (MAPE) values of 2.40 for training and 3.10 for testing. The decision tree for NO_x emissions had a depth of 11, with splits and leaves set at 9 and 3, respectively. This gave it a regulated structure that kept it from overfitting. Random Forest had 140 estimators, a maximum depth of 22, and "auto" features (Table 12). This showed that trees that were a little deeper and had more features made the model more accurate. This shows how NO_x reacts to many different things that work together, which ensembles are better at handling.

Due to a flawless training MSE of 0.00 and a training *R*-squared value of 1, the decision tree model precisely fits the training data. On unknown data, the testing MSE of 59.83 and a testing *R*-squared value of 0.9746 suggest, however, that the model shows some overfitting even if it shows strong performance. Finally, with *R*-squared values of 0.9869 for training and 0.9635 for testing, the Random Forest model shows a training MSE of 17.37 and a testing MSE of 85.89. Although the rather high testing MSE shows difficulties in generalization, represented in MAPE values of 1.54 for training and 4.41 for testing, these results show solid performance, stressing the model's constraints when forecasting NO_x emissions in fresh datasets.

5 Conclusions

In this study, experiments have been performed to evaluate the performance and emissions of a variable-speed diesel engine running with neat diesel and 20% biodiesel blends of *Manilkara zapota*, *Bauhinia variegata*, Karanja, and *Simarouba* oil. Comparisons of engine performance and emission from the test blends and neat diesel have been undertaken in this study. The following points can be concluded from the study:

• The oil contents of *M. zapota*, *Bauhinia variegata*, Karanja and *Simarouba* seeds are found to be 35%, 19%, 30% and 69%, respectively.

• The biodiesel fuel exhibits lower BP and BTE, higher BSFC, lower HC and CO emissions and higher NO_x emissions. The highest performance and lowest emissions from the engine for all considered biodiesels are achieved at an engine speed of around 3000 rpm.

• The average reductions in BP and BTE for MZME, BVME, KME and SME are approximately in the range of 3–6%, 7–11%, 5–9% and 8–12%, respectively.

• The average increases in BSFC for MZME, BVME, KME and SME are around 2%, 4%, 3% and 5%, respectively, compared to the diesel fuel.

• The average reductions in the emissions of HC and CO are in the range of 30%, 15–22%, 20–27% and 7–15% for MZME, BVME, KME and SME blends, respectively, and for NO_x emissions, the average increase are found to be 6%, 17%, 10 and 20%, respectively, compared to the diesel fuel.

• MZME shows comparative BP, BTE and BSFC, which are around 5.7% lower, 3.2% lower and 2.5% higher than those of diesel, respectively.

• MZME results in the maximum dip in CO and HC, which are found to be around 28.57% and 31% lower than those of the diesel fuel. Also, a marginal increase in NO_x (around 5.5% higher) is observed with MZME than with the diesel fuel.

• Performance and emission of *Manilkara zapota* biodiesel are observed to be superior to other biodiesel feedstocks, followed by *Bauhinia variegata* and Karanja; hence, *Manilkara zapota* seeds can be considered as a promising biodiesel source in India.

• The RSM-ML approach effectively identifies the optimal engine characteristics by combining the statistical modelling of parameter interactions with the predictive accuracy of machine learning, ensuring improved performance and efficiency.

• The Random Forest model demonstrates a lower MAPE compared to the other models, confirming its superiority in predictive modelling with respect to both generalization and accuracy balance.

In future studies, one can explore the areas related to engine performance and emission characteristics by mixing underutilized feedstocks of *Manilkara zapota*, *Bauhinia variegata*, Karanja, and *Simarouba* oil blends with nanoparticles and green fuel such as hydrogen at different blend ratios.

Conflicts of interest

The authors declare that there are no conflicts of interest or competing interests.

Data availability

The data that supports the findings of this study are available within the article.

Abbreviations

A/F	Air–fuel ratio
BTE	Brake thermal efficiency
BSFC	Brake specific fuel consumption
bTDC	Before top dead center
°C	Degree Celsius
CO	Carbon monoxide
CI	Compression ignition

CO_2	Carbon dioxide
CR	Compression ratio
DI	Direct injection
DOE	Design of experiments
HC	Hydrocarbon
ML	Machine learning
NO_x	Nitrogen oxide
RSM	Response surface methodology
RPM	Revolutions per minute
PPM	Parts per million
GHG	Greenhouse gases
MZME	*Manilkara zapota* methyl ester
BVME	*Bauhinia variegata* methyl ester
KME	Karanja methyl ester
SME	*Simarouba* methyl ester
LCV	Lower calorific value
kW	Kilowatt

Acknowledgements

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University, Saudi Arabia, for funding this work through the Research Group Program under Grant No. RGP. 2/366/46.

References

1 R. G. Mali, S. G. Mahajan and A. A. Mehta, Phcog Rev: Plant Review Rakta Kanchan (Bauhinia Variegata): Chemistry, Traditional and Medicinal Uses-a Review, *Pharmacogn. Rev.*, 2007, **1**(2), 14–19.

2 E. M. Shahid and Y. Jamal, Production of Biodiesel: A Technical Review, *Renewable Sustainable Energy Rev.*, 2011, **15**(9), 4732–4745, DOI: **10.1016/j.rser.2011.07.079**.

3 A. I. El-Seesy, A. M. A. Attia and H. M. El-Batsh, The Effect of Aluminum Oxide Nanoparticles Addition with Jojoba Methyl Ester-Diesel Fuel Blend on a Diesel Engine Performance, Combustion and Emission Characteristics, *Fuel*, 2018, **224**, 147–166, DOI: **10.1016/j.fuel.2018.03.076**.

4 I. A. Reşitoğlu, K. Altinişik and A. Keskin, The Pollutant Emissions from Diesel-Engine Vehicles and Exhaust Aftertreatment Systems, *Clean Technol. Environ. Policy*, 2015, **17**, 15–27.

5 A. I. EL-Seesy, M. Nour, H. Hassan, A. Elfasakhany, Z. He and M. A. Mujtaba, Diesel-Oxygenated Fuels Ternary Blends with Nano Additives in Compression Ignition Engine: A Step towards Cleaner Combustion and Green Environment, *Case Stud. Therm. Eng.*, 2021, **25**, 100911, DOI: **10.1016/j.csite.2021.100911**.

6 J. N. Nair, T. Nagadurga, V. D. Raju, H. Venu, S. Algburi, S. Kamangar, A. I. A. Arabi, A. Razak and N. Marakala, Impact of Fuel Additives on the Performance, Combustion and Emission Characteristics of Diesel Engine Charged by Waste Plastic Bio-Diesel, *Case Stud. Therm. Eng.*, 2025, 105755.

7 S. V. Ranganathan, S. L. Narasimhan and K. Muthukumar, An Overview of Enzymatic Production of Biodiesel, *Bioresour. Technol.*, 2008, **99**(10), 3975–3981, DOI: **10.1016/j.biortech.2007.04.060**.

8 P. Verma, M. P. Sharma and G. Dwivedi, Investigation of Metals and Antioxidants on Stability Characteristics of Biodiesel, *Mater. Today: Proc.*, 2015, **2**(4–5), 3196–3202, DOI: **10.1016/j.matpr.2015.07.114**.

9 S. Chattopadhyay and R. Sen, Fuel Properties, Engine Performance and Environmental Benefits of Biodiesel Produced by a Green Process, *Appl. Energy*, 2013, **105**, 319–326, DOI: **10.1016/j.apenergy.2013.01.003**.

10 J. Kjärstad and F. Johnsson, Resources and Future Supply of Oil, *Energy Policy*, 2009, **37**(2), 441–464, DOI: **10.1016/j.enpol.2008.09.056**.

11 A. S. Ramadhas, C. Muraleedharan and S. Jayaraj, Performance and Emission Evaluation of a Diesel Engine Fueled with Methyl Esters of Rubber Seed Oil, *Renewable Energy*, 2005, **30**(12), 1789–1800, DOI: **10.1016/j.renene.2005.01.009**.

12 J. vam Gerpen, Biodiesel Processing and Production, *Fuel Process. Technol.*, 2005, **86**(10), 1097–1107, DOI: **10.1016/j.fuproc.2004.11.005**.

13 G. Knothe, Dependence of Biodiesel Fuel Properties on the Structure of Fatty Acid Alkyl Esters, *Fuel Process. Technol.*, 2005, **86**(10), 1059–1070, DOI: **10.1016/j.fuproc.2004.11.002**.

14 A. E. Atabani, A. S. Silitonga, I. A. Badruddin, T. M. I. Mahlia, H. H. Masjuki and S. Mekhilef, A Comprehensive Review on Biodiesel as an Alternative Energy Resource and Its Characteristics, *Renewable Sustainable Energy Rev.*, 2012, **16**(4), 2070–2093, DOI: **10.1016/j.rser.2012.01.003**.

15 H. M. Mahmudul, F. Y. Hagos, R. Mamat, A. A. Adam, W. F. W. Ishak and R. P. Alenezi, Characterization and Performance of Biodiesel as an Alternative Fuel in Diesel Engines – A Review, *Renewable Sustainable Energy Rev.*, 2017, **72**, 497–509, DOI: **10.1016/j.rser.2017.01.001**.

16 A. K. Agarwal, Biofuels (Alcohols and Biodiesel) Applications as Fuels for Internal Combustion Engines, *Prog. Energy Combust. Sci.*, 2007, **33**(3), 233–271, DOI: **10.1016/j.pecs.2006.08.003**.

17 M. M. K. Bhuiya, M. G. Rasul, M. M. K. Khan, N. Ashwath and A. K. Azad, Prospects of 2nd Generation Biodiesel as a Sustainable Fuel—Part: 1 Selection of Feedstocks, Oil Extraction Techniques and Conversion Technologies, *Renewable Sustainable Energy Rev.*, 2016, **55**, 1109–1128, DOI: **10.1016/j.rser.2015.04.163**.

18 M. Mofijur, M. G. Rasul, J. Hyde, A. K. Azad, R. Mamat and M. M. K. Bhuiya, Role of Biofuel and Their Binary (Diesel–Biodiesel) and Ternary (Ethanol–Biodiesel–Diesel) Blends on Internal Combustion Engines Emission Reduction, *Renewable Sustainable Energy Rev.*, 2016, **53**, 265–278, DOI: **10.1016/j.rser.2015.08.046**.

19 S. Mitra, A. Ghose, N. Gujre, S. Senthilkumar, P. Borah, A. Paul and L. Rangan, A Review on Environmental and Socioeconomic Perspectives of Three Promising Biofuel Plants Jatropha Curcas, Pongamia Pinnata and Mesua

Ferrea, *Biomass Bioenergy*, 2021, **151**, 106173, DOI: 10.1016/j.biombioe.2021.106173.

20 P. Tamilselvan, N. Nallusamy and S. Rajkumar, A Comprehensive Review on Performance, Combustion and Emission Characteristics of Biodiesel Fuelled Diesel Engines, *Renewable Sustainable Energy Rev.*, 2017, **79**, 1134–1159, DOI: 10.1016/j.rser.2017.05.176.

21 P. Verma and M. P. Sharma, Review of Process Parameters for Biodiesel Production from Different Feedstocks, *Renewable Sustainable Energy Rev.*, 2016, **62**, 1063–1071, DOI: 10.1016/j.rser.2016.04.054.

22 A. Demirbas, Biofuels Sources, Biofuel Policy, Biofuel Economy and Global Biofuel Projections, *Energy Convers. Manage.*, 2008, **49**(8), 2106–2116, DOI: 10.1016/j.enconman.2008.02.020.

23 Y. G. Keneni and J. M. Marchetti, Oil extraction from plant seeds for biodiesel production, *AIMS Energy*, 2017, **5**(2), 316–340.

24 S. Rezania, B. Oryani, J. Park, B. Hashemi, K. K. Yadav, E. E. Kwon, H. Jin and J. Cho, Review on transesterification of non-edible sources for biodiesel production with a focus on economic aspects, fuel properties and by-product applications, *Energy Convers. Manage.*, 2019, **201**, 112155.

25 S. M. M. Hasnain, R. Chatterjee, P. Ranjan, G. Kumar, S. Sharma, A. Kumar, B. Salah and S. S. Ullah, Performance, Emission, and Spectroscopic Analysis of Diesel Engine Fuelled with Ternary Biofuel Blends, *Sustainability*, 2023, **15**(9), 7415.

26 R. Chatterjee, S. K. Mukherjee, B. Paul and S. Chattopadhyaya, Evaluation of Spectroscopic Analysis, Performance and Emissions of Enriched Jatropha and Madhuca Methyl Ester for Clean Environment, *Clean Technol. Environ. Policy*, 2022, **24**(7), 2295–2312.

27 A. Saravanan, M. Murugan, M. Sreenivasa Reddy and S. Parida, Performance and Emission Characteristics of Variable Compression Ratio CI Engine Fueled with Dual Biodiesel Blends of Rapeseed and Mahua, *Fuel*, 2020, **263**, 116751, DOI: 10.1016/j.fuel.2019.116751.

28 M. Hawi, A. Elwardany, S. Ookawara and M. Ahmed, Effect of Compression Ratio on Performance, Combustion and Emissions Characteristics of Compression Ignition Engine Fueled with Jojoba Methyl Ester, *Renewable Energy*, 2019, **141**, 632–645, DOI: 10.1016/j.renene.2019.04.041.

29 M. Arunkumar, M. Kannan and G. Murali, Experimental Studies on Engine Performance and Emission Characteristics Using Castor Biodiesel as Fuel in CI Engine, *Renewable Energy*, 2019, **131**, 737–744, DOI: 10.1016/j.renene.2018.07.096.

30 A. Uyumaz, Experimental Evaluation of Linseed Oil Biodiesel/Diesel Fuel Blends on Combustion, Performance and Emission Characteristics in a DI Diesel Engine, *Fuel*, 2020, **267**, 117150, DOI: 10.1016/j.fuel.2020.117150.

31 R. Sathish Kumar, K. Sureshkumar and R. Velraj, Combustion, Performance and Emission Characteristics of an Unmodified Diesel Engine Fueled with Manilkara Zapota Methyl Ester and Its Diesel Blends, *Appl. Therm. Eng.*, 2018, **139**, 196–202, DOI: 10.1016/j.applthermaleng.2018.04.107.

32 M. Vijayaragavan, B. Subramanian, S. Sudhakar and L. Natrayan, Effect of Induction on Exhaust Gas Recirculation and Hydrogen Gas in Compression Ignition Engine with Simarouba Oil in Dual Fuel Mode, *Int. J. Hydrogen Energy*, 2022, **47**(88), 37635–37647, DOI: 10.1016/j.ijhydene.2021.11.201.

33 R. Sathish Kumar, K. Sureshkumar and R. Velraj, Optimization of Biodiesel Production from Manilkara Zapota (L.) Seed Oil Using Taguchi Method, *Fuel*, 2015, **140**, 90–96, DOI: 10.1016/j.fuel.2014.09.103.

34 A. K. Agarwal and A. Dhar, Experimental Investigations of Performance, Emission and Combustion Characteristics of Karanja Oil Blends Fuelled DICI Engine, *Renewable Energy*, 2013, **52**, 283–291, DOI: 10.1016/j.renene.2012.10.015.

35 B. S. Chauhan, N. Kumar, H. M. Cho and H. C. Lim, A Study on the Performance and Emission of a Diesel Engine Fueled with Karanja Biodiesel and Its Blends, *Energy*, 2013, **56**, 1–7, DOI: 10.1016/j.energy.2013.03.083.

36 S. R. Mishra, M. K. Mohanty, N. Panigrahi and A. K. Pattanaik, Impact of Simarouba Glauca Biodiesel Blends as a Fuel on the Performance and Emission Analysis in an Unmodified DICI Engine, *Renew. Energy Focus*, 2018, **26**, 11–16, DOI: 10.1016/j.ref.2018.05.002.

37 S. Kandasamy and S. Sundararaj, Engine Performance and Emission Profile of Simarouba Glauca Biodiesel and Blends, *Int. J. Oil, Gas Coal Technol.*, 2020, **25**(2), 202, DOI: 10.1504/IJOGCT.2020.109444.

38 T. R. Praveenkumar, M. Sekar, R. R. Pasupuleti, B. Gavurová, G. Arun Kumar and M. Vignesh Kumar, Current Technologies for Plastic Waste Treatment for Energy Recovery, It's Effects on Poly Aromatic Hydrocarbons Emission and Recycling Strategies, *Fuel*, 2024, **357**, 129379, DOI: 10.1016/j.fuel.2023.129379.

39 Y. Wu, Y. Yuan, C. Xia, T. A. Alahmadi, S. A. Alharbi, M. Sekar and A. Pugazhendhi, Production of Waste Tyre Pyrolysis Oil as the Replacement for Fossil Fuel for Diesel Engines with Constant Hydrogen Injection via Air Intake Manifold, *Fuel*, 2024, **355**, 129458, DOI: 10.1016/j.fuel.2023.129458.

40 A. R. Shirneshan, M. Almassi, B. Ghobadian and G. H. Najafi, Investigating the Effects of Biodiesel from Waste Cooking Oil and Engine Operating Conditions on the Diesel Engine Performance by Response Surface Methodology, *Iran. J. Sci. Technol.*, 2014, **38**(M2), 289.

41 M. J. Eslami, B. Hosseinzadeh Samani, S. Rostami, R. Ebrahimi and A. Shirneshan, Investigating and Optimizing the Mixture of Hydrogen-Biodiesel and Nano-Additive on Emissions of the Engine Equipped with Exhaust Gas Recirculation, *Biofuels*, 2023, **14**(5), 473–484.

42 A. Zare, M. Babaie, A. Shirneshan, P. Verma, L. Yang, Z. D. Ristovski, R. J. Brown, T. A. Bodisco and S. Stevanovic, Hazardous Particles during Diesel Engine Cold-Start and Warm-up: Characterisation of Particulate Mass and Number under the Impact of Biofuel and Lubricating Oil, *J. Hazard. Mater.*, 2023, **460**, 132516.

43 A. Shirneshan, B. Kanberoglu and G. Gonca, Experimental Investigation and Parametric Modeling of the Effect of Alcohol Addition on the Performance and Emissions Characteristics of a Diesel Engine Fueled with Biodiesel-Diesel-Hydrogen Fuel Mixtures, *Fuel*, 2025, **381**, 133489.

44 A. Dewangan and A. Mallick, Ultrasonic-Assisted Production of Biodiesel from Manilkara Zapota (L.) Seed Oil, *Energy Sources, Part A*, 2017, 1–8, DOI: 10.1080/15567036.2017.1352631.

45 R. L. Patel and C. D. Sankhavara, Biodiesel Production from Karanja Oil and Its Use in Diesel Engine: A Review, *Renewable Sustainable Energy Rev.*, 2017, **71**, 464–474.

46 D. Kumar, V. Parcha, A. Maithani and I. Dhulia, Effect and Evaluation of Antihyperlipidemic Activity Guided Isolated Fraction from Total Methanol Extract of Bauhinia Variegata (Linn.) in Triton WR–1339 Induced Hyperlipidemic Rats, *Asian Pac. J. Trop. Dis.*, 2012, **2**, S909–S913.

47 N. Manojkumar, C. Muthukumaran and G. Sharmila, A Comprehensive Review on the Application of Response Surface Methodology for Optimization of Biodiesel Production Using Different Oil Sources, *J. King Saud Univ., Eng. Sci.*, 2022, **34**(3), 198–208.

48 N. T. Tefera, R. B. Nallamothu, G. A. Lakew and T. K. Kurse, Optimization of Biodiesel Production from Cottonseed Oil Using Response Surface Methodology and Artificial Neural Network Techniques, *Sci. Afr.*, 2025, **28**, e02665.

49 B. Chetia, S. Debbarma and F. Ahmed, Parametric Optimization of Variable Compression Ratio Diesel Engine Run on Waste Cooking Biodiesel with CeO2 Nanoparticles Using Response Surface Methodology (RSM) Technique, *Fuel*, 2025, **389**, 134544.

50 D. O. Onukwuli, C. Esonye, A. U. Ofoefule and R. Eyisi, Comparative Analysis of the Application of Artificial Neural Network-Genetic Algorithm and Response Surface Methods-Desirability Function for Predicting the Optimal Conditions for Biodiesel Synthesis from Chrysophyllum Albidum Seed Oil, *J. Taiwan Inst. Chem. Eng.*, 2021, **125**, 153–167.

51 G. Shanmugasundar, M. Vanitha, R. Čep, V. Kumar, K. Kalita and M. Ramachandran, A Comparative Study of Linear, Random Forest and Adaboost Regressions for Modeling Non-Traditional Machining, *Processes*, 2021, **9**(11), 2015.

52 K. R. Bukkarapu and A. Krishnasamy, Investigations on the applicability of machine learning algorithms to optimize biodiesel composition for improved engine fuel properties, *Int. J. Engine Res.*, 2024, **25**(7), 1299–1314.

53 L. Breiman, Random Forests, *Mach. Learn.*, 2001, **45**, 5–32.

54 J. Chen, X. Wang and F. Lei, Data-Driven Multinomial Random Forest: A New Random Forest Variant with Strong Consistency, *J. Big Data*, 2024, **11**(1), 34.

55 S. B. Kotsiantis, Decision Trees: A Recent Overview, *Artif. Intell. Rev.*, 2013, **39**, 261–283.

56 J. Zheng, J. Wang, Z. Zhao, D. Wang and Z. Huang, Effect of equivalence ratio on combustion and emissions of a dual-fuel natural gas engine ignited with diesel, *Appl. Therm. Eng.*, 2019, **146**, 738–751.

57 L. F. Chuah, A. R. A. Aziz, S. Yusup, A. Bokhari, J. J. Klemeš and M. Z. Abdullah, Performance and Emission of Diesel Engine Fuelled by Waste Cooking Oil Methyl Ester Derived from Palm Olein Using Hydrodynamic Cavitation, *Clean Technol. Environ. Policy*, 2015, **17**(8), 2229–2241, DOI: 10.1007/s10098-015-0957-2.

58 O. M. Ali, R. Mamat, N. R. Abdullah and A. A. Abdullah, Analysis of Blended Fuel Properties and Engine Performance with Palm Biodiesel–Diesel Blended Fuel, *Renewable Energy*, 2016, **86**, 59–67, DOI: 10.1016/j.renene.2015.07.103.

59 K. Nantha Gopal, A. Pal, S. Sharma, C. Samanchi, K. Sathyanarayanan and T. Elango, Investigation of Emissions and Combustion Characteristics of a CI Engine Fueled with Waste Cooking Oil Methyl Ester and Diesel Blends, *Alexandria Eng. J.*, 2014, **53**(2), 281–287, DOI: 10.1016/j.aej.2014.02.003.

60 D. H. Qi, H. Chen, L. M. Geng and Y. Z. Bian, Experimental Studies on the Combustion Characteristics and Performance of a Direct Injection Engine Fueled with Biodiesel/Diesel Blends, *Energy Convers. Manage.*, 2010, **51**(12), 2985–2992, DOI: 10.1016/j.enconman.2010.06.042.

61 J. Xue, T. E. Grift and A. C. Hansen, Effect of Biodiesel on Engine Performances and Emissions, *Renewable Sustainable Energy Rev.*, 2011, **15**(2), 1098–1116, DOI: 10.1016/j.rser.2010.11.016.

62 J. B. Hirkude and A. S. Padalkar, Performance and Emission Analysis of a Compression Ignition: Engine Operated on Waste Fried Oil Methyl Esters, *Appl. Energy*, 2012, **90**(1), 68–72.

63 A. Ahmad, A. K. Yadav and A. Singh, Enhancing Waste Cooking Oil Biodiesel Yield and Characteristics through Machine Learning, Response Surface Methodology, and Genetic Algorithms for Optimal Utilization in CI Engines, *Int. J. Green Energy*, 2024, **21**(6), 1345–1365.

64 S. M. Palash, M. A. Kalam, H. H. Masjuki, B. M. Masum, I. M. Rizwanul Fattah and M. Mofijur, Impacts of Biodiesel Combustion on NOx Emissions and Their Reduction Approaches, *Renewable Sustainable Energy Rev.*, 2013, **23**, 473–490, DOI: 10.1016/j.rser.2013.03.003.

65 A. Tamilvanan, B. Ashok, T. Mohanraj, P. Jayalakshmi, P. Dhamodharan and R. Sakthivel, Effect of Engine Operating Parameters in NOx Reduction, in *NOx Emission Control Technologies in Stationary and Automotive Internal Combustion Engines*, Elsevier, 2022, pp. 125–153.

66 N. R. Abdullah, Z. Michael, A. R. Asiah, A. J. Helmisyah and S. Buang, Effects of Palm Oil Methyl Ester (POME) on Fuel Consumption and Exhaust Emissions of Diesel Engine Operating with Blended Fuel (Fossil Fuel+ Jatropha Oil Methyl Ester (JOME)), *J. Teknol.*, 2015, **76**(5), 43–47.

67 T. Selvan and G. Nagarajan, Combustion and Emission Characteristics of a Diesel Engine Fuelled with Biodiesel Having Varying Saturated Fatty Acid Composition, *Int. J. Green Energy*, 2013, **10**(9), 952–965, DOI: 10.1080/15435075.2012.732157.

68 K. Raj Bukkarapu and A. Krishnasamy, Support Vector Regression Approach to Optimize the Biodiesel

Composition for Improved Engine Performance and Lower Exhaust Emissions, *Fuel*, 2023, **348**, 128604, DOI: 10.1016/j.fuel.2023.128604.

69 H. Li, W. Yang, D. Zhou and W. Yu, Numerical Study of the Effects of Biodiesel Unsaturation on Combustion and Emission Characteristics in Diesel Engine, *Appl. Therm. Eng.*, 2018, **137**, 310–318, DOI: 10.1016/j.applthermaleng.2018.03.066.

70 R. Manimaran, T. Mohanraj, M. Venkatesan, R. Ganesan and D. Balasubramanian, A Computational Technique for Prediction and Optimization of VCR Engine Performance and Emission Parameters Fuelled with Trichosanthes Cucumerina Biodiesel Using RSM with Desirability Function Approach, *Energy*, 2022, **254**, 124293.

71 P. Yang, H. Chen and Y. Liu, Application of Response Surface Methodology and Desirability Approach to Investigate and Optimize the Jet Pump in a Thermoacoustic Stirling Heat Engine, *Appl. Therm. Eng.*, 2017, **127**, 1005–1014.

72 U. A. Dodo, E. C. Ashigwuike and S. I. Abba, Machine Learning Models for Biomass Energy Content Prediction: A Correlation-Based Optimal Feature Selection Approach, *Bioresour. Technol. Rep.*, 2022, **19**, 101167.