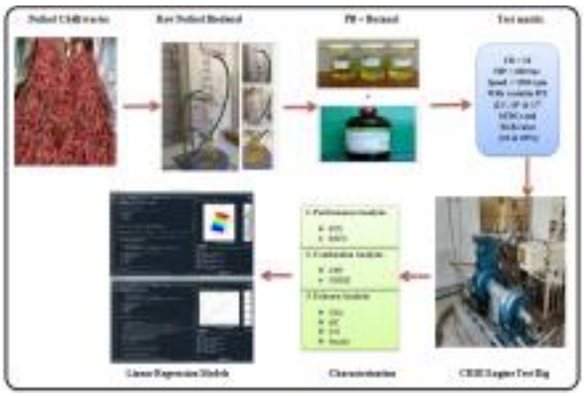
Global NEST Journal, Vol 27, No 6, 05532

A statistical prediction of thermal efficacy improvement and exhausts reduction of CRDI engine energized with ternary blends using linear regression machine learning model

A. N. Phani Deepthi[a]*, J Prabakaran[b], A P Sathiyagnanam[c] and Srinivasa Prasad Sanaka[d]
[a]Department of Mechanical Engineering, Annamalai University, Chidambaram, Tamil Nadu, India.
[b]Department of Mechanical Engineering, ACGCET (Deputed from Annamalai University), Karaikudi, India
[c]Department of Mechanical Engineering, Govt. College of Engineering (Deputed from Annamalai University), Salem, India
[d]Department of Mechanical Engineering, V R Siddhartha Engineering College, Kanuru, Vijayawada, A.P, India
Received: 15/11/2023, Accepted: 26/03/2025, Available online: 05/05/2025
*to whom all correspondence should be addressed: e-mail: qinyan@cqjtu.edu.cn
https://doi.org/10.30955/gnj.05532

Graphical abstract



Abstract

The intensifying environmental and energy-related issues have led to a significant surge in the need for discovering alternative fuel sources. Additionally, the labor and time involved in identifying the appropriate fuel for the engine being examined is considerable. In order to mitigate this workload, the utilization of statistical analysis might be employed in the pursuit of identifying suitable fuel blends. The present work utilizes a machine learning algorithm (MLA) to predict the emission and performance characteristics of a common rail direct injection engine fueled by waste pedicel biodiesel and Butanol 20 dual fuel. The research focuses on examining ignition, efficiency, and exhaust traits of various fuel blends with varying Injection timing (21°, 19° & 17° bTDC), with Exhaust Gas Recirculation of 10% and 20% also employed during experimentation. After performing a series of investigations PBD20Bu is used to draw maximum performance with minimized emissions with other test samples. This study employed the linear regression model. As a result, the R^2 values of 0.98, 0.977, 0.975, 0.98, and 0.947, respectively, the findings suggest that the total accuracy of the predictions exceeds 94%. Shows that the linear regression model created exhibits a high level of accuracy in predicting the levels of NOx, CO, HC, smoke, and BTE.

Keywords: Pedicel biodiesel, n-butanol, CRDI engine, EGR, and linear machine learning model.

1. Introduction

Diesel engines are thermally efficient than gasoline engines with an additional advantage of reducing global warming. Deficiency in fossil sources probes towards usage of renewable sources thus stood as potential substitute by showing better performance in addition to cleaner emissions on diesel engines (Atmanli et,al.,(2018) & Ashok, *et al.* (2019)). Vegetable oils from non-edible feedstock like rapeseed oil, karanja seed oil, jatropha oil (M Karabektas *et al.* (2009)), castor oil, beef tallow oil, recycled frying oils (Kumar MS *et al.* (2005)). Practical approach reveals that the continuous supply of this feedstock around the year can only ensure said efficiencies with lowered emissions based on conclusions drawn from experimental procedures (Reddy N *et al.* (2016)). Plant based Bio-wastes conversion starts from collection of feedstock, drying, grinding, extraction, transesterification. Optimization of these processes can lead to closure behavior as conventional diesel. Some of the conventional methods for the extraction of bio- diesels are Maceration, Decoction, sonication, counter current extraction, percolation, supercritical Extraction, percolation, hot continuous extraction also called Soxhlet Extraction. Among these Soxhlet is classical and effective method (Knothe *et al.* (2005) & Ghobadian B *et al.* (2008)). This plant extract from Soxhlet extractor which is a mixture of solute and solvent is separated conventionally by steam distillation process. Due to heat losses and fitting losses only 40% of the solvent in 150min at 70°C can be retrieved leading to non-economic approach (Luque de C *et al.* (2010)). So Rotary evaporator is employed by the later researchers for the same volume of mixture where about 65-70% in 35min at

70°C of the Solvent can be retrieved with the separation of pure Bio-oil (Dutta, R *et al.* (2014)). One of the functional parameter is selection of solvent. Some of the solvents in usage are n-Hexane, Iso-octane, Diethyl ether, Iso-propanol, Toluene, Benzene and chloroform (Yadessa *et al.* (2021)). Among all n-Hexane is considered to be effective above 1.3% higher in optimum conditions (Ali, *et al.* (2015)) tested with 5 different solvents for extraction of putranja seeds. This Bio-oil is characterized using GC-MS and FTIR analysis. Based on the aromatic compounds present in the oil the Bio-oil is then transesterified to Bio-diesel in order to remove Free fatty acid content and separate byproducts like esters and glycerol using a standard procedure with the aid of an acid catalyst (Sayyar, S *et al.* (2009)). This method is considered as most viable technique intended to reduce Bio-oil viscosity and conversion of methyl esters. Physicochemical properties (Shiu, P.J *et al.* (2010)) like Kinematic Viscosity, fuel Density, Flash point, fire point, Iodine value, saponification value, pH value, heating value of Bio-diesel are evaluated using ASTM standard methods. This research mainly focuses on synthesis of Bio-oil and the classical procedure involved right from the collection of feed stock from Industry Guntur Mirchi Yard to preparation of Bio-diesel. Bio-oil and bio-diesel are characterised using ASTM standard methods, followed by comparison with conventional diesel fuel and existing bio-diesel. This Pedicel Biodiesel is the used to evaluate the performance, combustion and emission characteristics to assess the extent of suitability in par with the existing Bio-diesels. Present study is on completely new plant based feedstock which is originated from Capsicum annuum var longum called Guntur sannam. This species is most hot pepper species also called cayenne with scoville heat unit range is 35,000-40,000 SHU with 10% moisture content (Junaid A *et al.* (2014)). For a long shelf life of this fruit the stem, only part where bacteria can grow is removed. India is the largest producer of species producing 23 lakh ton stood first in the world in exporting Dry chilli. In this 38% is from Andhra Pradesh shares maximum export due its humid climatic conditions. In this 11lakh metric tons production is from Guntur Mirchi Yard, Guntur (Arora R *et al.* (2011)). From 11 Lakh ton we get about 2.26 Lakh metric ton of pedicel waste dumped. The research carried out showed that digital technology has advanced and how aquaponics, the fifth industrial revolution, has changed production patterns. Thus, machine learning-based predictive maintenance detects industrial equipment issues to decrease downtime and improve maintenance schedules. Machine-learning-powered quality control systems detect production faults and assure product quality. Machine learning algorithms increase supply chain logistics, inventory management, and efficiency. Process optimisation finds production data patterns using machine learning to optimize efficiency and energy utilisation. Machine learning-enabled autonomous robots and intelligent decision-making systems simplify complex jobs for humans. The industrial environment is becoming more connected, intelligent, and data-driven thanks to aquaponics well ((R Surendran *et al.* (2023)). Conducted an exhaustive examination of the machine-learning methods utilised to forecast the exhaust characteristics and efficiency of IC engines. Artificial neural networks (ANNs) were utilised with considerable frequency. The research emphasized the potential of machine learning and deep learning to assist in the determination of internal combustion engine characteristic forecasts. Additionally, the limitations of artificial neural networks were acknowledged, including unexplained network behaviour and an increase in computational complexity due to an accuracy dependence on the number of concealed layers (Karunamurthy *et al.* (2023)). A machine-learning model based on artificial neural networks (ANN) was designed to forecast the density and viscosity of petroleum fuels that incorporate oxygenated chemical classes, including alcohol. The programme utilises the molecular structure of the compounds as an input riable. The training of the model involved the utilisation of a dataset consisting of 164 individual pure chemicals as well as 14 composite mixtures with known compositions. The neural network model demonstrated appropriate predictions, as evidenced by the density correlation factors of 0.99 and the viscosity correlation factors of 0.98 (AlNazr *et al.* (2023)). This research uses intelligent industrial defect detection uses a residual network with IIFD-SOIR. Modelling feature extraction, categorization, and signal representation. CWT preprocesses vibration data. ResNet v2 Inception retrieves key characteristics. Its left-module shortcut connections make it superior than other DL models. The better IIFD-SOIR scored 99.6% gearboxes and 99.64% bearings. The model fared better in simulations (R Surendran *et al.* (2022)). The author has created a method that uses machine learning methods to fully automate the process of optimizing software for internal combustion engines. The system included a variety of learning strategies, including a Super Learner model, ensemble-based learners, optimization algorithms, and active learning. In order to address the multi-objective problem, a simplification was undertaken, transforming it into a single-objective challenge, which subsequently yielded a commendable merit score (Mohan and Badra, (2023)).

1.1. Aim and objective

Scientific goals include finding eco-friendly fuels and improving operating characteristics. Internal combustion engine testing is complex, expensive, and time-consuming. Thus, engine and fuel research communities and the industry are actively seeking reliable computational approaches to suit these objectives. Experimental datasets and machine learning can estimate diesel, biodiesel and ethanol mix CRDI engine characteristics. As with ternary fuel mixes, machine learning models can anticipate test factors to increase engine combustion and performance and minimise emissions. Algorithms may analyse engine sensor data and change fuel mixes to improve performance and minimise emissions. Machine learning algorithms can estimate engine emissions from a variety of gasoline mixes by considering characteristics like fuel blend, load fuel usage, and others.

1.2. Motivation and scope of the Work

The suitability of a novel biodiesel in combustion-ignition engines is widely recognized owing to its elevated cetane number. The primary aim of this study is to evaluate the characteristics of a CRDI diesel engine by utilizing different blends of waste pedicel biodiesel and diesel, specifically under full load conditions, without making any modifications to the engine. Biodiesel derived from a novel capsicum waste is combined with conventional diesel and n-butanol at different concentrations of PBD20 and PBD20Bu. The experimental investigation focused on analysing the cylinder pressure, heat release rate, brake thermal efficiency (BTE), and engine tailpipe emissions of a CRDI diesel engine. The engine was fueled with ternary and binary blends of pedicel biodiesel and n-butanol. The experimental outcomes were then compared and discussed with the baseline diesel fuel. The current study is founded around the utilisation of the Linear Regression (LR) approach to effectively simulate the engine's performance and emissions when operated with biofuel-diesel mixtures. The objective of this study is to precisely simulate the impact of various engine loads, injection parameters, and blending ratios on the efficiency of energy transformation and emission profile of a stationary CRDI engine. Additionally, a prediction model based on linear regression is constructed utilising well-established indicators of statistics.

2. Materials and methods

2.1. Feedstock collection bio-oil extraction process

Feedstock is collected from Industry and sundried at 34˚C. This is powdered in grinders. The powdered feedstock is taken in batches in Soxhlet extractor to get extract. A 500ml round bottom flask filled with 320ml n-hexane is kept in a heating mantle.120gm of pedicel powder is taken in a thimble loaded in the main chamber of the extractor. Water is circulated in the reflux condenser to the top of it. Temperature of the heating mantle is set to the boiling point of the solvent about 69˚ C. The vapours of the solvent will pass over the biomass and cooled in the condenser. Converted liquid vapour will drip back into the main chamber and emptied due to siphoning action which is shown in the figure. The extract collected will be filled in round bottom flask, placed in a water bath of a Rotary evaporator which is rotated at constant speed of 130rpm. The evaporation system is maintained at 400-600 mm of Hg. This rotation brings the flask to negative pressure and initiates evaporation. About 70 % of the Solvent is thus retrieved from the extract and pure oil will be remained in the flask. This process is carried with series of extractors continuously in batches to get pure Bio-oil (at Accurate Labs, Vijayawada). This series of operations were performed with different solute and solvent mixtures 1:4,1:6, 1:8 and 1:10 ratios and observed that bio-oil yield is maximum at 1:8 as tabulated using the equation given below

$$\left(\text{\% of bio oil yield}\right) = \frac{\text{Chili pedicel powder bio oil}}{\text{Chili pedicel powder}\left(\text{gm}\right)} \times 100$$

2.2. Bio-oil characterization of GC-MS analysis

Characterization of Bio-oil is done using GC-MS analysis to known the FFA content composition and types of hydrocarbons present (L.K.Velayutham *et al.* (2015), & Shah, Z *et al.* (2016)). Test setup is shown in the figure (at IIT, Madras). A sample volume of 1.0gm bio-oil in chloroform was injected through a split mode, with 15:1 split ratio and split flow of 18mL/min at 11.367psi. Fatty acid composition from GC-MS test is tabulated shown. From the GC-MS analysis, 18 FFA compounds were found listed in the table. Among them 59% of unsaturated and 41% of saturated fatty acids were present in this pedicel oil.

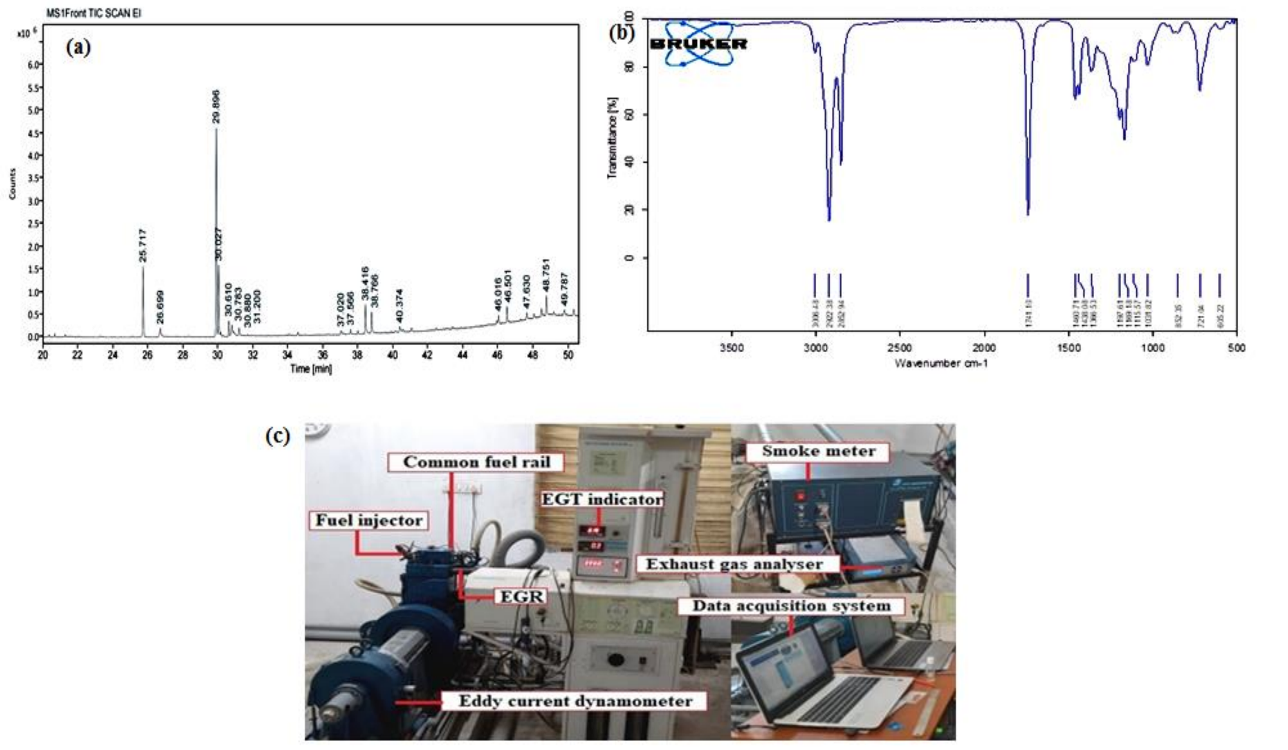


Figure 1. (a) GC-MS Chromatograph for 9, 12-Octadecadienoic acid of Pedicel Bio-oil (b) FTIR Spectra in the wavenumber range 1000-2000 cm^{-1} and (c) Photographic view of CRDI VCR Engine Test Setup

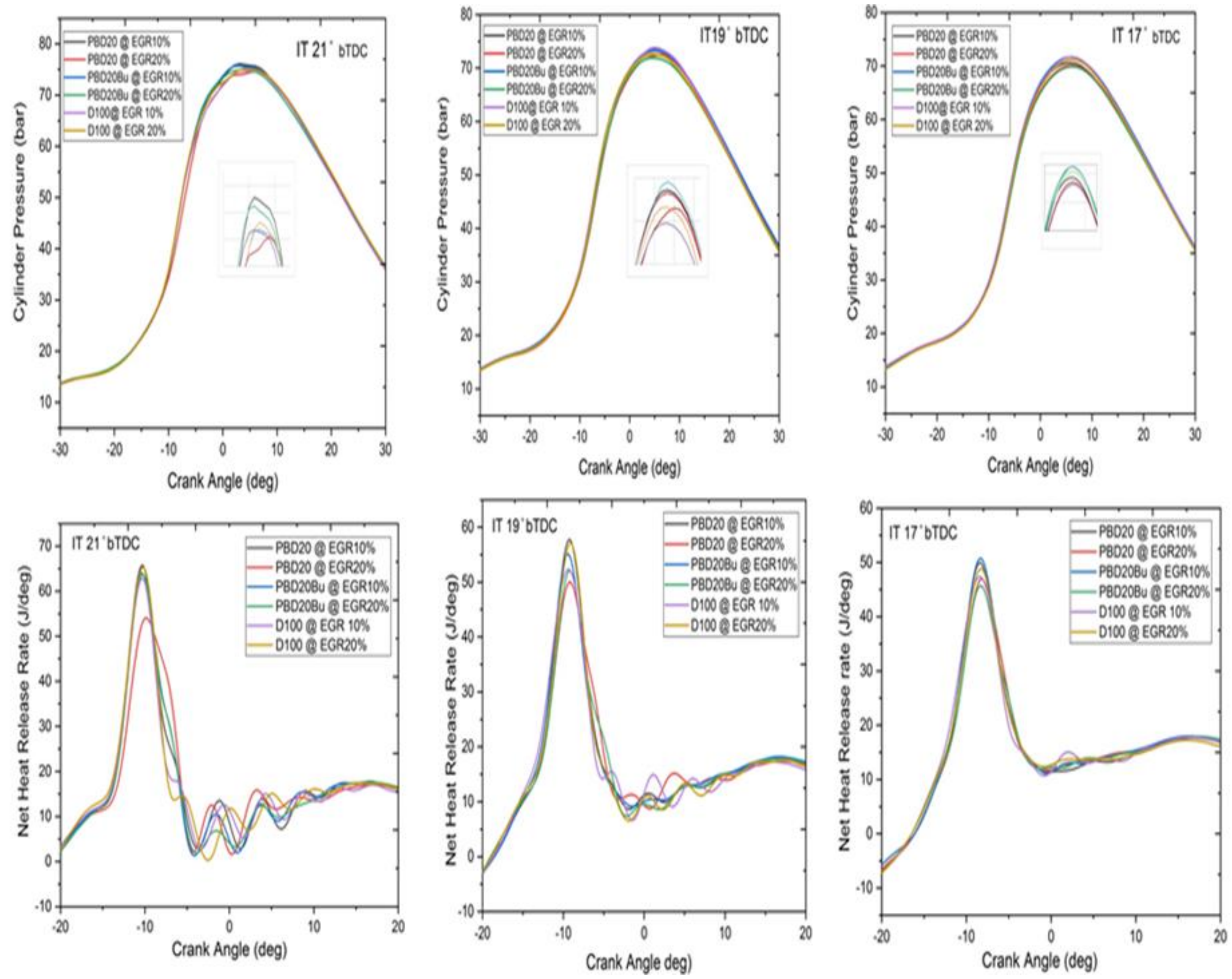


Figure 2. variations of Cylinder pressure and Net HRR (vs) Crank Angle at different FITs

2.3. Fourier transform infrared (FTIR)

Fourier Transform Infrared (FTIR) spectroscopy analyzer is used to know polymer, organic and inorganic content of the Bio-oil (Sutrisno, B *et al.* (2018), & Subramanian, N *et al.* (2015)) specifically it's a method to analyze nature of the chemical bonds based on stretching and bending of Infrared radiation. Perkin Elmer spectrum IR analyzer (at Accurate Labs, Vijayawada) used to carry out this study.

The FTIR spectrum is recorded from 400 to 4000cm^{-1} in the transmission mode for pedicel Bio-oil. Sample Spectra from 1000 to 2000cm^{-1}wave number range is shown in the **Figure 1(B)**. The presence of oxygenated functional groups indicates the oxygen percentage in the bio-oil. The presence of hydrocarbon group indicates that this is the potential substitute as the alternative source of energy. After characterization of pure Bio-oil, transesterification is carried.

2.4. Transesterification

Bio-oil is mixed with methyl alcohol in the ratio 16:1 then by adding 1% H_2SO_4 mixture is heated at 65˚ C for 30min, acid value decreases to a range less than 4 KOH/g (Sharma, V *et al.* (2014), & Degfe, T *et al.* (2019)). 750ml Bio-oil and 400ml CH_3OH including NaOH are subjected to stirring at 1200rpm for a period of 45min. At the end of this process Methyl esters, glycerol, and traces of NaOH are identified. Traces were removed by purification using HCL and water. The obtained Pedicel Bio-diesel is 96.4% pure biodiesel. Some properties of obtained pedicel bio-diesel were evaluated using ASTM standards like kinematic viscosity is 7.57cst, Cetane Index of 58.75, Flash point 14˚C, fire point 22˚C, pour point of below minus 18 ˚C and Gross calorific value of 10557Kcal/kg. These were carried out in ITA Labs, Chennai, and Tamilnadu. Some of the properties Saponification is 155.24, Iodine values 114.34 and pH of 5.46 are also evaluated at Accurate Labs, Vijayawada. The properties were tabulated for different types of test samples as shown.

2.5. Experimental setup

The experiment is carried out on Kirloskar single cylinder VCR with CRDI equipped diesel engine. Real experimental setup is shown in the **Figure 1(C)** used to study performance and emission characteristics with pedicel biodiesel blends. The engine is interlinked to open ECU(Nira i7r swedeh)with solenoid injector driver, Compression pressure sensor(PCB piezotronics, USA), Kubler Crank angle sensor resolution 1 resolution, speed 5500 rpm with TDC pulse, Yokogawa fuel flow transmitter with a range of 0-500mm WC, Pressure transmitter range 250WC and a high speed data acquisition device of 16bit, 250KS/s. The data monitoring, reporting, entry and data logging are performed with engines developed by Apex Innovations. Techno Mech-10 Eddy current dynamometer, water cooled used as loading unit. In order to regulate the exhaust gas circulation level, the ECU is also connected to the water-cooled EGR unit, which has a vacuum device and pneumatic equipment valve. AVL DI GAS 444N (five gas analyser) used to measure the exhaust emission content from the engine exhaust within response time of ≤15sec.

Table 1. Properties of Test Samples

Parameter	Diesel	PBD	ASTM Standards
Density (kg/m^3)	836	877	D 4052
Kinematic Viscosity(cst)	2.71	7.57	D 445
Cetane Index	52	58.75	D 4737
Flash Point (°C)	75	14	D 92
Net CV (Kcal/kg)	10355	10187	D240

Table 2. Specifications of Test Engine

Particulates	Engine Details
Make and Model	Kirloskar, TV1
Number of cylinders	One
Stroke	Four
Bore x stroke	87.5 X 110mm
Swept volume	661cc
Compression ratio	17.5
Rated output	3.5 kW at 1500rpm
Rated speed	1500 rpm
Cooling system	Water cooled
Injection timing, CA bTDC	23°
Injection pressure	600bar

2.6. Test methodology

Diesel is procured from Indian Oil. Pedicel Biodiesel blends with 3 different volumetric proportions 10%, 20%, and 30% were mixed using magnetic stirrer. Experimentation is initiated with pure diesel D100; a blend of this diesel fuel with 10%Pedicel Biodiesel (PBD10)by volume; a blend of this diesel fuel with 20%Pedicel Biodiesel (PBD20)by volume; a blend of this diesel fuel with 10%Pedicel Biodiesel (PBD30)by volume were homogeneously mixed using magnetic stirrer. n-butanol is further used as an additive fraction 5ml (PBD20Bu) for the best blend after conducting series of investigations which tends to be a bit beneficiary. Performance and emission characteristics of CRDI diesel engine using prepared test fuels. The after treatment system EGR (Exhaust Gas Recirculation 10% and 20% were also employed during experimentation. Experiment is carried out by varying load conditions from no load to full load and Injection Timings 23° BTDC, 21° bTDC, 19° bTDC and 17° bTDC at constant injection pressure of 500 bar and constant speed of 1500rpm to know the best combinations.

2.7. Uncertainty analysis

For each independent parameter mean, standard deviation and standard error are calculated. For this purpose a set of 12 readings were selected to ensure the uncertainty and errors while carrying out experimental investigations (Knothe *et al.* (2005)). This analysis helps in finding out errors and uncertainities due to incorrect readings, observations, environment conditions, selection of instrument and calibration.

Total Uncertainty during experimentation is given by the following equation

$$= \sqrt{\begin{bmatrix} (NO)^2 + (CO)^2 + (HC)^2 + (S)^2 + (L)^2 + (T)^2 \\ +(B.P)^2 + (B.T.E)^2 + (F.C)^2 + (P)^2 + (\theta)^2 + (M)^2 \end{bmatrix}}.$$

The uncertainty values of individual parameters were listed in the **Table 3**.

2.8. Machine Learning algorithm (MLA):

The study of machine learning is a domain of AI that is dedicated to the advancement of algorithms as well as statistical models capable of acquiring knowledge from data and generating predictions. Linear regression is a supervised machine-learning algorithm that falls under the category of machine learning. It operates through learning from datasets with labels, establishing mappings between observations, and optimizing linear coefficients. These functions may then be utilised for making predictions on fresh datasets (Karthikeyan Subramanian *et al.* ((2024)). First and foremost, it is necessary to establish a comprehensive understanding of supervised machine learning algorithms. Supervised learning is a machine learning paradigm in which an algorithm acquires knowledge from a set of labeled data.

2.9. Linear regression model (LR):

The goal of linear regression, a supervised machine learning process, is to find out the linear connection between a dependent variable and one or more independent features. When there is just one independent function, we call it unitary linear regression; when there are multiple features, we call it multiplex linear regression. The purpose of the approach is to locate the optimal linear equation for making predictions about the variable being dependent on known values of the variables that are autonomous. A straight line, depicting the relationship between the dependent and independent variables, is provided by the function. The slant of the line represents the proportional change in the dependent variable for each one-unit shift in the variables that are independent. Regression is widely recognized as a crucial task in the field of supervised learning. Within a regression dataset, there exist records containing both X and Y values. These values are utilised to acquire knowledge about a function. Consequently, if one desires to forecast the value of Y based on an unknown X, this acquired function can be employed. In the context of regression analysis, it is necessary to determine the value of the dependent variable, denoted as Y. In the context of regression analysis, there is a need for a function that can accurately predict continuous values of Y while considering X as a set of distinct variables.

3. Outcomes and discussions

Tests were carried at constant speed of 1500 rpm and by varying 0 to 100% load. At every trial were conducted by varying injection timing 17°, 19° and 21° bTDC with 10% and 20% EGR rates without any modifications in the engine. With this the effect of blends and blends with additive n-butanol on combustion, performance, and emission characteristics are studied.

Table 3. Showing uncertainities for various measuring parameters

Measurement	Uncertainty (%)	Measurement	Uncertainty (%)
Nitrous oxides(NO)	0.2	Brake power (B.P)	0.2
Carbon Monoxide(CO)	0.1	Brake Thermal Efficiency (B.T.E)	0.8
Hydrocarbons(HC)	0.3	Fuel Consumption (F.C)	0.6
Speed(S)	0.2	Cylinder Pressure(CP)	0.8
Load(L)	0.8	Crank Angle(θ)	0.1
Time(T)	0.6	Manometer (M)	0.2

The accuracy of the experimental investigations is found using error analysis and the uncertainty percentage is calculated as ±1.705%.

3.1. Discussion on combustion parameters

3.1.1. Pressure (vs) crank angle

Cylinder pressure is plotted with respect to crank angle at 100%load for various blends at IT 21° BTDC, IT19° BTDC and IT17° BTDC as shown in the **Figure 2**. The peak pressure inside the cylinder obtained forPBD20 at 10%EGR is about 76.07bar at IT21° BTDC. A bit higher compared to pure diesel at similar condition of EGR, 73.83bar retarding 2° CA. This is due to the concentration of pedicel biodiesel whose cetane number is higher than diesel leading the shorter ignition delay, so by adding additive n-butanol the peak pressure reduces and helps in maintaining enough delay for complete combustion (Nanthagopal *et al.* (2018)). The PCP of methyl acetate mixtures decreases as EGR rates escalate. This occurs because the inert gases included in the redirected exhaust gases serve as heat sinks by absorbing the energy released during combustion. The maximum pressure inside the cylinder is lowered (Ashok, *et al.* (2019)).

3.1.2. NHRR (vs) crank angle

The peak pressure inside the cylinder also leads to Higher Heat release rates. **Figure 2** shows Net Heat release rates for different blends at various Injection timings. Initially till the fuel injection starts a sudden dip in the heat release rate is observed while with a negative value of net heat release rate as here heating and vaporization of fuel droplets occurs (Ashok, *et al.* (2019)). As the combustion starts a sudden rise in the Heat release rate is noted. The first peak of Net HRR is observed as 68.21 J/deg at IT21° BTDC for the D100 at 20%EGR and 59.56 J/deg at IT 19° BTDC. At IT17° BTDC for the blend PBD20Bu at 10%EGR first peak is at 51.96 J/deg. This clearly shows the impact of Injection timing on the heat release rates. Higher heat release rates for D100 corresponds to pre-mixed combustion as the fuel will be burned rapidly and this causes rapid increase in pressure and temperature in few degrees of crank angle. Another reason for this may be the accumulation fuel droplets inside the chamber. Second peak in HRR is observed which is much lower than the first peak in duration of about 27° CA the HRR gradually decreases (El-Seesy *et al.* (2020)). At the end of the combustion it slows down and here about 15 to 20% of remaining heat is released. As the injection timing is advanced increase in ignition delay period is observed.

3.2. Performance analysis

Brake thermal efficiency is the amount of heat that is converted from the supply into braking power. **Figure 3** depicts the Brake Thermal efficiency at 100% engine load for various blends with change in injection timings. It is observed that at IT 17° bTDC, PBD20 shows Brake thermal efficiency nearer to D100 with 10%EGR rate. similarly at 20% EGR rate PBD20 shows closer to D100with 20% EGR.For the Blend PBD20Bu it is observed that at IT 21° bTDC the BTE is very nearest to neat diesel D100 at 20%EGR as higher cetane numbers and oxygen content of the blend PBD20 exhibits complete combustion with no heat loss to the evaporation of fuel droplets and this is the reason behind showing better BTE.

The higher cetane value of blend shorter ignition delay, which decreases period of combustion and thereby the amount heat shifted to the fragments of the engine, reduces during combustion process. So due to the reduction in heat loss the BTE increases (Subramanian, K., *et al.* (2022)). The effect of EGR on BTE at 100%load for is test fuels are also compared in the **Figure 3**. The BTE of engine decrease with increase in EGR rate and also with the addition of n-butanl a slight decrease in BTE is observed as EGR replaces fresh charge coming into the combustion chamber thereby lowers combustion temperature (Kumar, S, *et al.* (2019)). It is noted that 3.04% reduction for he blend PBD20Bu at IT 19° BTDC. It is clear that from the graph when IT 19° BTDC is retarded for a particular blend there is increase in BTE.

3.3. Exhaust emissions analysis

3.3.1. NOx

NOx Emissions purely depends on residence time and temperature of the mixture inside the combustion chamber (Rajesh Kumar et, al.(2016)). **Figure 4** shows NOx emissions from the CRDI engine at 100% load for all test fuels for varied Injection timings. Studies were also done on EGR rate and impact of injection timing on the emissions. It is observed that PBD20 blend exhibits lower emissions of NOx than D100 at EGR10%. D100 NOx emissions at 100%

load, IT 17° bTDC, EGR 10% is 2083ppm, for the blend PBD20 it is 1788ppm, for the blend PBD20Bu it is 1467ppm. This clearly shows a gradual reduction of NOx due to addition of PBD and n-butanol. Similar trend is observed for 20%EGR. Further increase in EGR% shows reduction of NOx for all test fuels at similar conditions. This impact of injection timing clearly depicts that NOx reduces gradually with the retardation in injection timing (Kumar, S, *et al.* (2019)). PBD20 at 10%EGR emits 2384ppm at IT 21° bTDC, 2134ppm at IT 19° bTDC, 1788ppm at IT 17° bTDC. As EGR replaces fresh charge, clearly depicts that both injection timing and EGR govern NOx reduction.

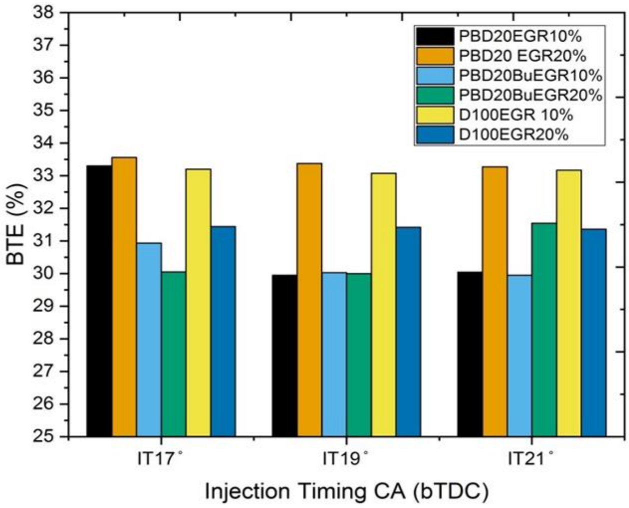


Figure 3. Variations of BTE at different Injection timings

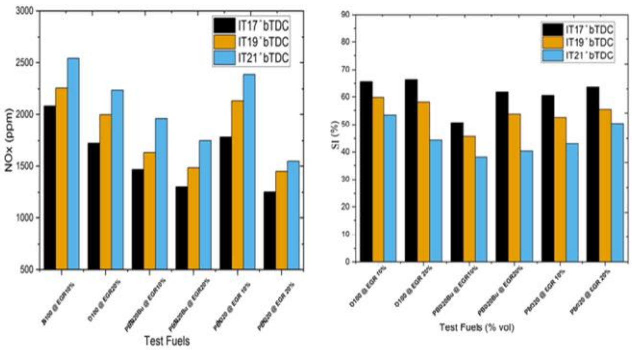


Figure 4. NOx and Smoke for various blends at varying Injection timings

3.3.2. Smoke opacity

When light in the visible range from a source is transmitted through a certain path length of the exhaust gases, smoke opacity is the fraction of that light which prevents from reaching the observer of smoke meter (Kumar, S, *et al.* (2019)). Smoke Opacity for with PBD at 100% load at various Injection timings for all test fuels are shown in the **Figure 4**. From the graphical representation it is clear that at 100% load smoke released is lower at IT21° bTDC for the test fuel PBD20Bu compared with diesel due to high vaporization rate. Also at 100% loads HRR and flame temperature is high which leads to lower levels of smoke in exhaust. PBD20 at 10% and 20% EGR shows less smoke when compared to D100 at same conditions as highly volatility reduces smoke at higher loads and this may be the reason (Subramanian, K., *et al.* (2022)). When Injection timing is retarded there is a gradual increase in smoke content is observed due to the diffusion combustion phase resulted by reduction in fuel air mixing rate due to later injection is responsible for this trend (K.Subramanian et, al, (2023)).

3.3.3. HC

Hydrocarbon emissions depend on the extent of complete combustion in the cylinder. **Figure 5** shows HC emissions from CRDI at 100%load for various test fuels at varied injection timings. It has been noted that D100 at 10% EGR emits 68ppm at IT 21BTDC at similar condition PBD20 emits 70ppm and for PBD20Bu it is 57. This clearly depicts that initially a slight increase in HC is observed but on the addition of n-butanol HC ultimately decreased. This is due to the enhancement in oxygen content in the blend which ensures complete combustion. But the increase in EGR % a slight increase in HC emission is observed (Subramanian, K., *et al.* (2022)). At 10%EGR, at IT 17 bTDC engine emits 71 ppm whereas at 20% EGR rate it emits about 85ppm. Similar trends are observed for PBD20 and PBD20Bu. As EGR probes to form a large quenching zone which results incomplete combustion (Nanthagopal *et al.* (2018)). As the IT is retarded initially emissions are noted to be slightly higher further retardation leads to decrease in HC emissions for D100 with 10%EGR rate. The HC emissions were observed to be lower for the test fuel PBD20Bu at 10% EGR and IT21 bTDC.

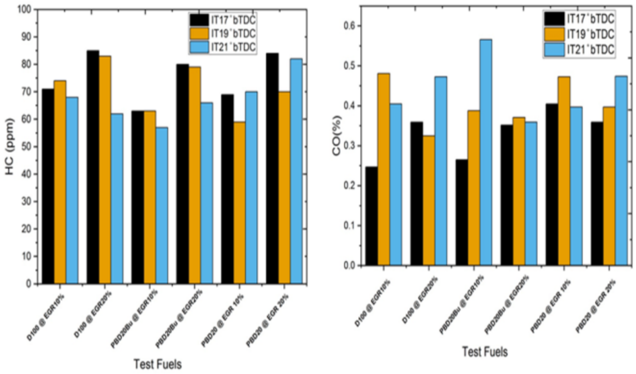
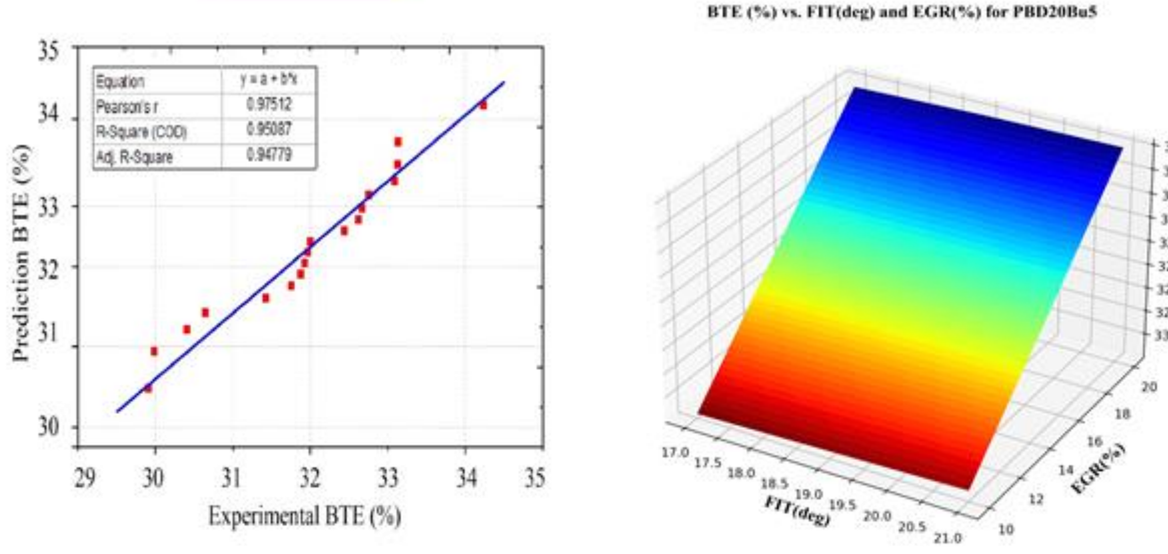


Figure 5. HC and CO for various blends at varying Injection timings

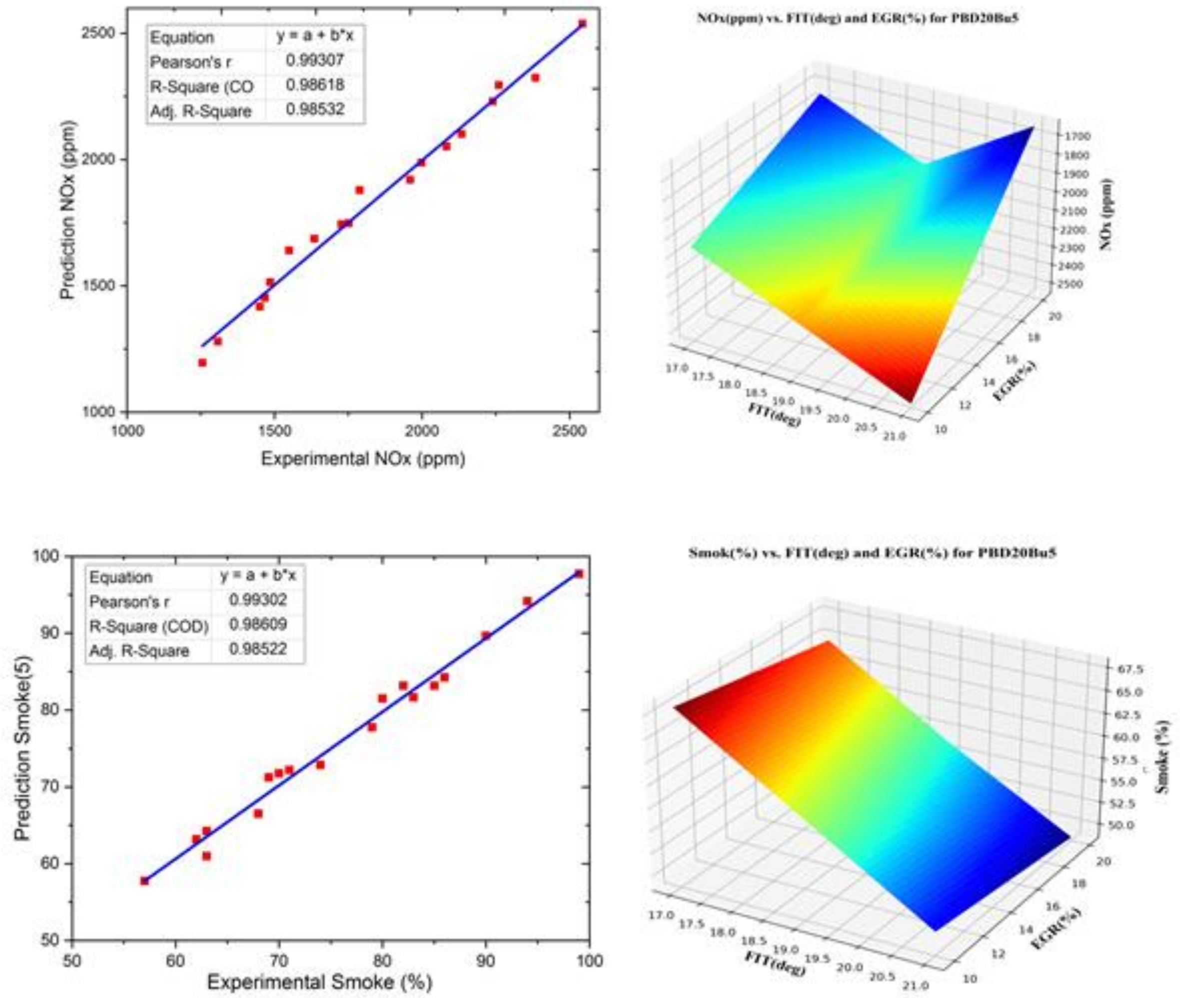
3.3.4. CO emissions

CO emissions from CRDI with all test fuels are shown in the **Figure 5**. CO emission was found to be high for PBD20Bu at 10%EGR and IT21BTDC due to high viscosity which improves heat of vaporization. Also viscosity increases spray penetration which tends to enhance wall deposition (B Rajesh Kumar *et al.* (2015)). D100 at 10% EGR emits 24.7% which is noted to be low CO emission, at similar condition PBD20 emits a bit higher CO noted as 40% but with the additive n-butanol it has been reduced to 26.5%, nearer to D100.Retardation in IT shows reduction in CO. At 21° bTDC PBD20Bu at 10% EGR shows 66.8%, with retardation it reduces to 38.8% and further retardation to IT17 BTDC 26.5%emissions. EGR enhancement leads to increase in emissions for test fuels D100 andPBD20Bu as this process dilutes fresh charge coming into the cylinder results in poor oxidation (El-Seesy *et al.* (2020)).

(a) Prediction of Efficiency model:



(b) Prediction of Emission models:



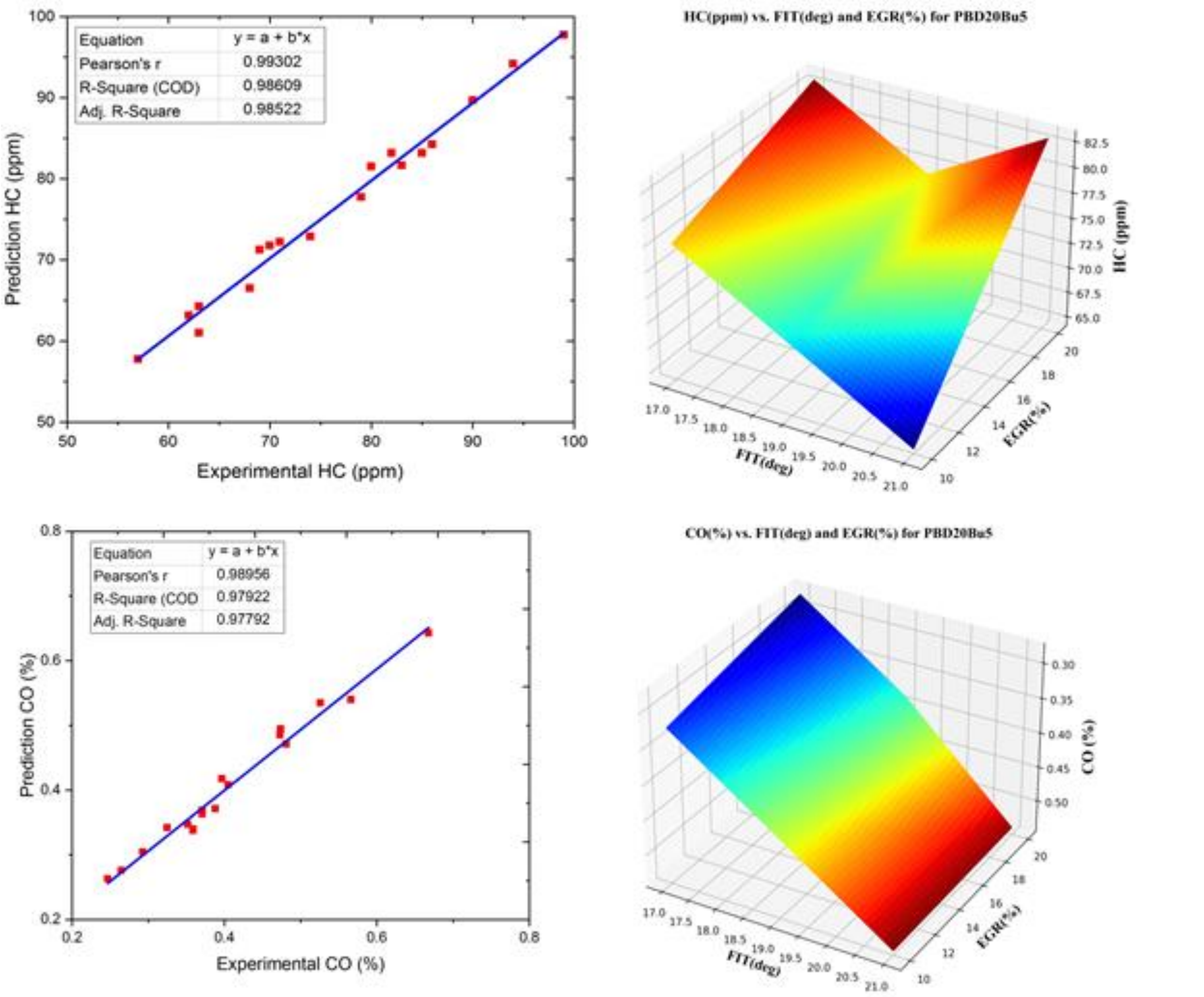


Figure 7 displays the linear regression correlations

4. Evaluation of machine learning modeling

Data obtained from physical experiments is typically stored using the software application Excel. The predictor variables in this study include load, injection time, fuel fraction, and EGR. On the other hand, the response variables being studied are carbon monoxide, hydrogen, smoke, nitrogen oxide, and brake thermal efficiency. Approximately 298 distinct experiments are conducted, wherein input parameters are systematically altered and the corresponding outcomes are meticulously documented.

The data undergoes preprocessing and is subsequently trained using a random forest algorithm. The execution of a sensitivity analysis comes after the partitioning of the dataset into training and testing sets in an 80:20 ratio. To mitigate any bias in the selection of occurrences, a randomized division process is employed. The obtained results are then used to identify the ideal values for the input attributes. **Figure 6** illustrates the sequential procedures undertaken to determine the most favorable outcome. The dataset undergoes a thorough examination to identify any instances of missing values. 15 data points are missing for both the fuel fraction and injection time features. The mean value of each feature is computed and subsequently employed to populate the missing data. Subsequently, an analysis of correlation is conducted in order to ascertain the degree of linearity across the predictor and responder variables Specifically, the calculation of the correlation between the input variables and each output variable is performed using Pearson's correlation coefficient. Values or correlations range from -1 to +1, denoting the extent of a positive to negative connection. Linearity is observed when the values closely approximate either +1 or -1. **Figure 7** displays the variables' correlation.

The correlation between the input and output features has a significantly low value, falling within the range of -1 and +1. This observation provides justification for employing methodologies that are appropriate for non-linear models. Therefore, in order to mitigate such irregularities, the data is subjected to normalization by utilising min-max normalization, resulting in a standardized range between 0 and 1. To mitigate anomalies arising from disparate numeric ranges, the training dataset is subjected to normalization. The squared correlation (R^2) can be computed by using Equ (1)

$$\mathbf{R}^2 = 1 - \left(\frac{\sum_{\mathbf{i}=1}^{\mathbf{n}} (\mathbf{t_i} - \mathbf{o_i})^2}{\sum_{\mathbf{i}=1}^{\mathbf{n}} (\mathbf{o_i})^2} \right) \quad (1)$$

The machine-learning models were predicted using the following steps:

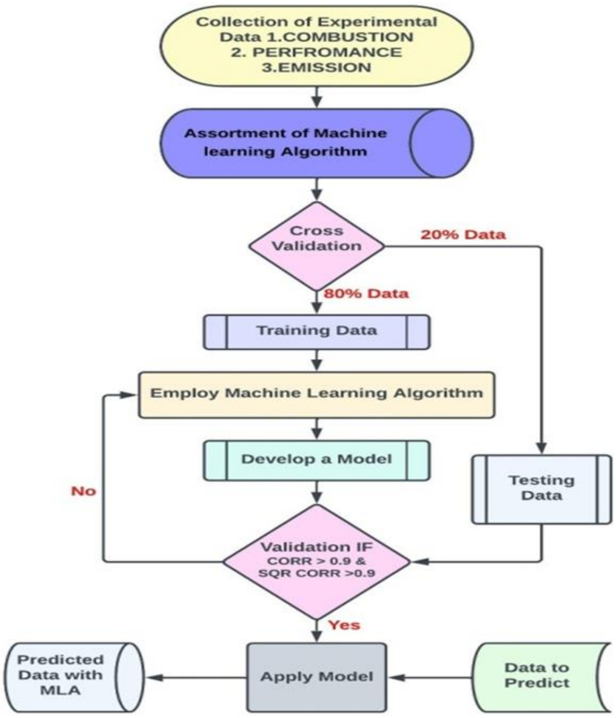


Figure 6. The Flowchart of Linear regression model

1. Data collection

The goal of data collection is to gather data, engine parameters, and performance metrics for model training. Engine parameters include FIT, EGR, and fuel composition (e.g., biodiesel, ethanol). Performance metrics, or target variables, include thermal efficiency, fuel consumption, and exhaust emissions (CO, NO_x, CO_2, particulates). The data should be structured in a tabular format (e.g., CSV or Excel), with rows representing test samples and columns for both features and target variables, ensuring easy analysis and model training.

2. Data processing

Data processing plays a vital role in preparing the dataset for training a linear regression model. The first step involves addressing missing data, which can be handled by either imputing values (using mean, median, or mode) or removing rows and columns with significant gaps. Categorical variables**,** like fuel types, need to be converted into numerical values using methods such as one-hot encoding or label encoding to ensure compatibility with the regression model.

3. Data split (Training and Testing)

Train-Test Split: Split the dataset into a training set (usually 70-80%) and a testing set (usually 20-30%). This allows the model to be trained on one set and evaluated on an unseen set, helping assess its generalization ability.

4. Model training

Train the linear regression model using the training dataset.

Fit a linear regression model to predict the target variable using the selected features. The linear regression formula is:

$$Y = \beta 0 + \beta 1 X1 + \beta 2 X2 + \ldots + \beta n Xn + \varsigma$$

Where Y is the target variable (e.g., thermal efficiency or emissions), X_1, X_2,…,Xn are the input features, and β_0,β_1,…,βn are the coefficients and ε - error.

5. Model evaluation

This involves assessing the performance of a trained model on unseen data to ensure it generalizes well beyond the training set. This is typically done by testing the model on a separate testing set and measuring its accuracy using metrics like R-squared (R^2), and These metrics help determine how closely the model's predictions align with actual outcomes. Effective evaluation helps identify whether the model is overfitting or underperforming, guiding adjustments for improved accuracy and reliability.

5. Conclusions

Combustion characteristics Peak pressure inside the cylinder, Mean Gas Temperature and Net heat release rate first tends to increases for PBD20 compared to D100 at 100%load but later with the addition of n-butanol (PBD20Bu) all said characteristics tends to reduce due to the premixed combustion.

An enhancement in EGR from 10% to 20% and retardation of Injection timing the Peak pressure, and NHRR reduced.

Brake Thermal efficiency is higher for the test fuel PBD20 at 20%EGR rate with lower NOx emissions at 100% load when compared to D100 at IT 17° bTDC, however with retardation in injection timing the brake thermal efficiency tends to increase for this blend. When EGR rate is enhanced BTE increases.

NOx emissions were noted to be lower for test fuel PBD20 at 20%EGR rate compared to D100 at same EGR rate. Increase in EGR rate and retardation in injection timing shows reduction in NOx.

HC emissions are found to be lower for the test fuel PBD20Bu at EGR 10% also found to be bit closer to PBD20 at similar conditions. Increase in EGR rate shows enhancement in HC emissions. Retardation in Injection timing shows slight increase in HC emissions for the test fuel PBD20Bu.

CO emissions are lower for PBD20Bu at EGR 10% and also seem to be closer to D100. Enhancing EGR rate reduces CO%. Retardation in Injection timing also reduces the CO% for the test fuels PBD20 and PBD20Bu.

Smoke is found to be low for the blend PBD20Bu an EGR 10%. With an increase in EGR rate retardation in Injection timing Smoke opacity tends to increase.

In statistical terms, it can be observed that linear regression outperforms forecasting ability due to its ability to achieve smaller mistakes and higher efficiency. The primary advantage of the linear regression (LR)-based prediction model is its ability to build a closed-form statistical link between controlled input factors and output (response) factors. The findings suggest that the total accuracy of the predictions exceeds 94%, respectively.

Hence it can be concluded that the Pedicel Bio-diesel with n-buanol as additive can be a potential alternative and conclusions drawn from the both examinations.

6. Limitations

The study has several limitations, including reliance on high-quality data, which may be incomplete or noisy, potentially leading to biased predictions. The linear regression model assumes a simple linear relationship, which may not capture the complexity of real-world engine behaviours. It also focuses on a specific set of ternary fuel blends, limiting the generalizability of the results. The controlled test conditions may not fully reflect real-world driving scenarios, and variations in fuel quality, especially biofuels, could impact engine performance. Additionally, the findings may not apply to other engine types beyond CRDI engines. These factors suggest a need for further research and advanced modelling to enhance the accuracy and scope of the study.

7. Nomenclature

ASTM	American Society for Test Materials
PB20	20% Pedicel Biodiesel + 80% Diesel
BTE	Brake Thermal Efficiency
bTDC	Before Top Dead Center
CR	Compression Ratio
IT	Injection Timing
EGT	Exhaust Gas Temperature
HSD	High Speed Diesel
CA	Crank Angle in degree
NOx	Oxide of Nitrogen (ppm)
HC	Hydrocarbon (ppm)
CO	Carbonmonooxide (%)
SO	Smoke Opacity (%)
Bu	Butanol

References

Ali, M. A., Al-hattab, T. A. and Al-hydary, I. A. (2015). Extraction of date palm seed oil (phoenix dactylifera) by Soxhlet apparatus. *International Journal of Advanced Engineering and Technology* **8**, 261–271.

Arora R, Gill N S, Chauhan G, Rana A C. (2011). An Overview about Versatile Molecule Capsaicin *International Journal of Pharmaceutical Sciences and Drug Research*; **3**(4): 280–286 ISSN 0975-248X.

Ashok, B., K. Nanthagopal, M. Senthil Kumar, A. Ramasamy, D. Patel, S. Balasubramanian and S. Balakrishnan. (2019). "An investigation on CI engine characteristics using pork lard methyl ester at various injection pressures and injection timings. *International Journal of Green Energy*, **16**(11): 834–846.

Atmanli A. and Yilmaz N. (2018). A comparative analysis of n-butanol/diesel and 1- pentanol/diesel blends in a compression ignition engine, *Fuel* **234**, 161e169, https://doi.org/10.1016/j.fuel.2018.07.015.

Balaji Mohan, Jihad Badra. (2023). An automated machine learning framework for piston engine optimization, *Applications in Energy and Combustion Science*, **Volume 13**, 100106, ISSN 2666–352X, https://doi.org/10.1016/j.jaecs.2022.100106.

Degfe, T. A., Mamo, T. T. and Mekonnen, Y. S. (2019). Optimized biodiesel production from waste cooking oil (WCO) using calcium oxide (CaO) nano catalyst. *Scientific Reports* **9**, 18982.

Dutta, R., Sarkar, U. and Mukherjee, A. (2014). Extraction of oil from Crotalaria Juncea seeds in a modifed Soxhlet apparatus: physical and chemical characterization of a prospective bio-fuel. *Fuel* **116**, 794–802.

El-Seesy, A. I., He Z., Hassan H. and Balasubramanian D. (2020). "Improvement of combustion and emission characteristics of a diesel engine working with diesel/jojoba oil blends and butanol additive." *Fuel* **279**: 118433.

Ghobadian B, Rahimi H, Nikbakht AM, Najafi G and Yusaf TF. (2008). Diesel engine performance and exhaust emission analysis using waste cooking biodiesel fuel with an artificial neural network. *Renewable Energy*; **34**(4):976–82.

Hussain A. AlNazr, Nabeel Ahmad, Usama Ahmed, Balaji Mohan, Abdul Gani Abdul Jameel. (2023). Predicting physical properties of oxygenated gasoline and diesel range fuels using machine learning, *Alexandria Engineering Journal*, **Volume 76**, Pages 193–219, ISSN 1110–0168, https://doi.org/10.1016/j.aej.2023.06.037.

Junaid Ahmad a, Suzana Yusup a, Awais Bokhari a, Ruzaimah Nik Mohammad Kamil. (2014). Energy Conversion and Management, *Study of Rubber Biodiesel*, **Volume 78**, pp. 266–275.

Karabektas M., Hosoz M. (2009). Performance, and emission characteristics of a diesel engine using isobutanol-diesel fuel blends, *Renew. Energy* **34**, 1554–1559, https://doi.org/10.1016/j.renene.2008.11.003.

Karthikeyan Subramanian, Sathiyagnanam Amudhavalli Paramasivam, Damodharan Dillikannan, Ravikumar Jayabal. (2023). ”Effect of pilot fuel injection strategies and EGR on a CRDI diesel engine powered by simmondsia chinensis seed biodiesel-methyl acetate blend”, *Sustainable Energy Technologies and Assessments*, **Volume 58**, 103345,

Karthikeyan Subramanian, Sathiyagnanam Amudhavalli Paramasivam, Damodharan Dillikannan. (2024). Machine learning predictions on the output parameters of CRDI engines fueled with ternary blend: *Chemical Industry & Chemical Engineering Quarterly*.

Karunamurthy, K., Janvekar, A.A., Palaniappan, P.L. *et al.* (2023). Prediction of IC engine performance and emission parameters using machine learning: A review. *Journal of Thermal Analysis and Calorimetry* **148**, 3155–3177. https://doi.org/10.1007/s10973-022-11896-2

Knothe *et al.* (2005). “Kinematic viscosity of biodiesel fuel components and related compounds. Influence of compound structure and comparison to petro diesel fuel components”. *Fuel* **84**, 1059–1065.

Kumar MS, Kerihuel A, Bellettre J, Tazerout M. (2005). Experimental investigations on the use of preheated animal fat as fuel in a compression ignition engine. *Renewable Energy*, **30**:1443–56.

Kumar, S., P. Dinesha and M. A. Rosen. (2019). "Effect of injection pressure on the combustion, performance and emission characteristics of a biodiesel engine with cerium oxide nanoparticle additive." *Energy* **185**: 1163–1173.

Luque de Castro, M. D. and Priego-Capote, F. (2010). Soxhlet extraction: past and present panacea. *Journal of Chromatography A* **1217**, 2383–2389.

Nanthagopal, K., Ashok, B., Saravanan, B., Mathew, S. and Chandra, S. (2018). Efect of next generation higher alcohols and Calophyllum inophyllum methyl ester blends in diesel engine. *Journal of Cleaner Production* **180**, 50–63.

Rajesh Kumar B. *et al*. (2015). Effect of EGR on performance and emissions of a constant speed di diesel engine fueled with pentanol/diesel blends, *Fuel* **160**, 217e226, https://doi.org/ 10.1016/ j.fuel.2015.07.089.

Rajesh Kumar B., Saravanan S. (2016). Use of higher alcohol biofuels in diesel engines: a review, *Renewable and Sustainable Energy Reviews* **60**, 84e115, https://doi.org/ 10.1016/j.rser.2016.01.085.

Reddy N, Ramesh A. (2006). Parametric studies for improving the performance of a jatropha oil-fuelled compression ignition engine. *Renewable Energy*; **31**:1994–2016.

Sayyar, S., Abidin, Z. Z., Yunus, R. and Muhammad, A. Extraction of oil from jatropha seeds-optimization and kinetics. *American Journal*.

Shah, Z., Cataluna Veses, R. and Silva, R. D. (2016). GC-MS and FTIR analysis of bio-oil obtained from freshwater algae (spirogyra) collected from Freshwater. *International Journal Environmental Agriculture Research*. **2**, 2454–1850.

Sharma, V., Garg, G. and Alam, A. (2014). Extraction, and characterization of industrially valuable oil from Eruca sativa (L.) Mill through FTIR and GC-MS analysis. *American Journal of Biological Chemistry*. 2, 23–28.

Shiu, P. J., Gunawan, S., Hsieh, W. H., Kasim, N. S. and Ju, Y. H. (2010). Biodiesel production from rice bran by a two-step in-situ process. *Bioresource Technology* **101**, 984–989.

Subramanian, K., *et al.* (2022). Prediction of thermal performance and exhaust emissions of a diesel engine fuelled with Simmondsia Chinensis oil methyl ester using ANN. *International Journal of Ambient Energy*, p. 1–14

Subramanian, N., Mahendradas, D. K., Kasirajan, R. and Sahadevan, R. (2015). Bio-oil separation from potential non-edible urban waste source Putranjiva roxburghii. *Separation Science and Technology* **50**, 2066–2074

Surendran, R., Khalaf, O.I., Tavera Romero, C.A. (2022). Deep learning based intelligent industrial fault diagnosis model. *Computers, Materials & Continua*, **70**(3), 6323–6338.

Sutrisno, B. and Hidayat, A. (2018). Pyrolysis of palm empty fruit bunch: Yields and analysis of bio-oil. *MATEC Web of Conferences* **154**, 1–4.

Velayutham L.K., Damodaran K. (2015). “An Economic Analysis of Chillies Production In Guntur District of Andhra Pradesh”. *International Journal of Research in Economics & Social Sciences*, **Volume 5**, Issue 9.

Venkatraman, M. and Surendran, R. (2023). "Industrial 5.0 Aquaponics System Using Machine Learning Techniques," *7th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, **2023**, pp. 835-842.

Yadessa Gonfa Keneni, Legesse Adane Bahiru and Jorge Mario. (2021). Marchetti Effects of Different Extraction Solvents on Oil Extracted from Jatropha Seeds and the Potential of Seed Residues as a Heat Provider, *Bio Energy Research*, **volume 14**, pages1207–1222.