STM JOURNALS

International Journal of
Mechanical Dynamics and Systems Analysis

Volume 3, Issue 1, 2025
January–June
DOI (Journal): 10.37591/IJMDSA

https://journals.stmjournals.com/ijmdsa

Research IJMDSA

Machine Learning Approach to Predict the Performability and Emissions of Diesel Engine Fueled with Doped Biodiesel Blend

Kiran Dinkar Chaudhari[1,*], Kailas Dhanraj Deore[1], Kshamta Mathur[2]

Abstract

Enhancing the performability and emission characteristics of diesel engines has been a difficult task in light of growing concerns about global warming and other negative effects, as diesel accounts for 70% of global energy demand. In this study, engine performance and exhaust emissions for various fuel blends were thoroughly evaluated using machine learning techniques to predict engine emission and performance behavior. We focused on biodiesel blend and nanoparticle additive concentration to gain valuable insights into optimizing engine performance and emissions for alternative fuel blends, advancing knowledge in this field. Four machine learning algorithms—decision tree, random forest, linear regression, and XGBoost—were employed to predict engine performance (brake-specific fuel consumption (BSFC), brake thermal efficiency (BTE), exhaust gas temperature (EGT)) and emission parameters (CO, CO_2, HC, NOx, smoke) using data from experiments with various biodiesel blends (B10, B20, B30) derived from waste cooking oil and diesel, incorporating different calcium oxide nanoparticle concentrations. The models were trained and tested using a 70/30 split of a dataset comprising engine load, additive concentration, and biodiesel blend ratio as input variables. Model performance was evaluated using R^2, mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). With regression coefficient (R2), MSE, MAE, MAPE, and RMSE values of 0.845, 255.7, 5.66, 0.1645, and 7.49 percent, respectively, in overall prediction, the XGBoost model performed better in engine performance prediction. This study demonstrates the effectiveness of XGBoost in optimizing biodiesel blends and nanoparticle additives for enhanced engine performance and reduced emissions, offering a more efficient alternative to traditional experimental methods. Future research could explore more advanced machine learning techniques and validate the model with a larger dataset.

***Author for Correspondence**
Kiran Dinkar Chaudhari
E-mail: kiran.chaudhari022@nmims.in

[1]Student, Department of Data Science, Shri Vile Parle Kelavani Mandal's Narsee Monjee Institute of Management Studies (SVKM's NMIMS), Mukesh Patel School of Technology Management and Engineering, Mumbai, Maharashtra, India
[2]Assistant Professor, Department of Data Science, Shri Vile Parle Kelavani Mandal's Narsee Monjee Institute of Management Studies (SVKM's NMIMS), Mukesh Patel School of Technology Management and Engineering, Mumbai, Maharashtra, India

Received Date: May 20, 2025
Accepted Date: June 07, 2025
Published Date: June 17, 2025

Citation: Kiran Dinkar Chaudhari, Kailas Dhanraj Deore, Kshamta Mathur. Machine Learning Approach to Predict the Performability and Emissions of Diesel Engine Fueled with Doped Biodiesel Blend. International Journal of Mechanical Dynamics and Systems Analysis. 2025; 3(1): 1–12p.

Keywords: Biodiesel, emission, engine performance, machine learning, nanoparticles

INTRODUCTION

Diesel engines are extensively used in heavy-duty transportation and power generation because of their significant benefits in terms of cost-effectiveness, fuel efficacy, and high efficiency [1]. However, by releasing greenhouse gases like CO, CO_2, and NOx, these engines play a major role in environmental degradation and global warming [2]. Alternative fuels have been created as viable alternatives to conventional fossil fuels in order to

allay these worries [3]. The demand for diesel engine substitutes has accelerated due to the gradual depletion of fossil fuel supplies, unreliable oil rates, and strict exhaust regulations [4]. Owing to its similar fuel qualities and lower emissions of smoke and greenhouse gases, biodiesel has become a viable alternative [5].

Owing to their higher oxygen content and the absence of harmful elements such as sulfur or minerals, vegetable, and animal fats result in lower emissions. The calorie value of biodiesel made from animal and vegetable fats is comparable to that of conventional diesel [6]. Although conventional diesel fuels do not have this problem, the high concentration of oxygen atoms in biofuels might cause corrosion [7]. In addition, biofuels use gasoline more specifically for brakes than traditional diesel fuels. Inherent oxygen atoms can enhance combustion properties and exhaust emissions, but they can also negatively affect stability and thermal efficiency. An increase in NOx emissions, the most dangerous exhaust gas pollutant arising from diesel engines, is the primary disadvantage of biofuels. Therefore, enhancing the quality of biodiesel and diesel is essential to encourage the use of alternative fuels. The use of fuel additives is a widely accepted strategy to modify fuel properties and curate the disadvantages of using biodiesel blends in engines. Several fuel additives based on metal oxides and carbon derivatives such as graphene, alcohols, and water have been used by researchers have successfully tried over the past decade [8]. Most of the research has focused on the use of oxides of metals, such as cerium, titanium, copper, silver, aluminum, and iron, which are detrimental to the environment to varying degrees. However, the use of oxides such as CaO has rarely been documented in the literature.

Therefore, further investigation into the use of CaO in biodiesel blends may yield valuable insights for developing alternative fuels that are both more sustainable and efficient. This research aligns with the ongoing pursuit of environmentally friendly energy solutions. The paper begins with a review of the existing literature, followed by a detailed description of the data, variables, and procedures used for fuel blend preparation and engine setup. The research methodology is then outlined, leading into a discussion of the experimental results and their implications. Finally, the study concludes with a summary of key findings and suggestions for future research directions.

LITERATURE REVIEW

Global Energy Scenario

According to the World Bank Data, the global population has grown by more than a billion between 2010 and 2021, spurring urbanization of rural areas and increasing demands for social mobility, as shown in Figure 1 [9].

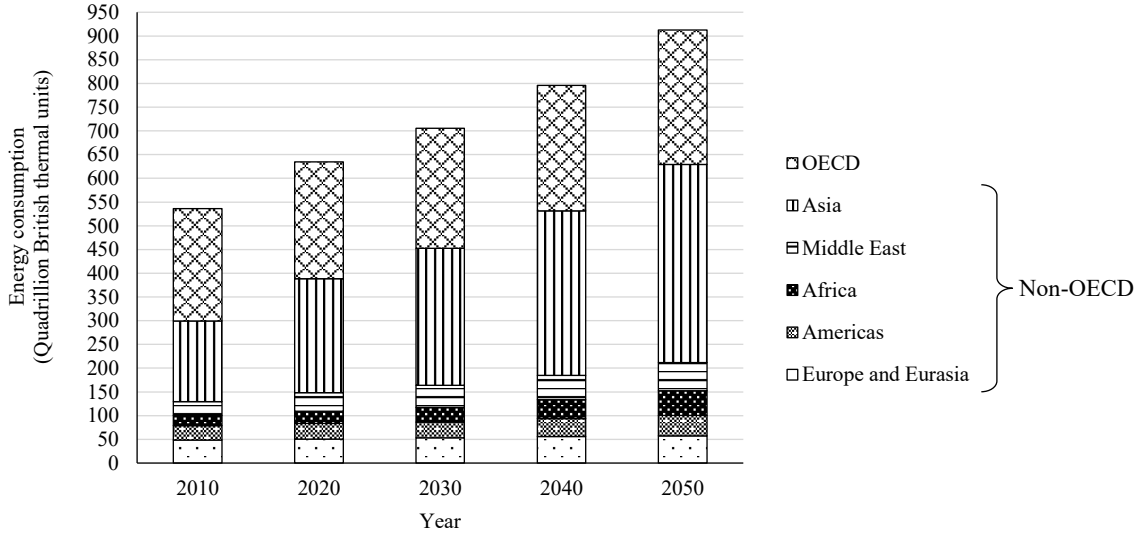

Figure 1. Projected global energy consumption till 2050.

This trend has led to a higher need for automotive vehicles, resulting in increased fossil fuel consumption and consequently increasing greenhouse gas emissions, exacerbating climate change and global warming. In response, many governments have established automotive industry, setting minimum emission standards to balance energy efficiency with fuel consumption rates. Three primary approaches have emerged to enhance engine performance, energy efficiency, and emission reduction: modifying combustion systems, utilizing renewable fuels, such as biofuels and biodiesel, and employing exhaust gas treatments.

Previous research indicates that while conventional diesel emits more carbon monoxide (CO) and unburnt hydrocarbons (HC) than biodiesel, emissions of nitrogen oxides (NOx) remain higher, particularly under heavy engine loads with biodiesel [10]. Historically, NOx emissions have been managed by adjusting fuel injection timing. Currently, blending biodiesel with conventional diesel and incorporating additives has demonstrated promising results in further improving the emission standards and efficiency.

Fuel Additives Used in Engines

Experimental studies using oxygenated fuels such as methanol, ethanol, and n-butanol as additives with diesel have shown improvements in emission reduction with minimal impact on engine performance [11]. Similar results have been documented when oxygenated fuels were blended with biodiesel. In recent years, nanoparticles have been explored for their impactful role in various engineering fields, from enhancing heat transfer to reducing friction and wear, and even in optical and cooling applications. As nanoadditives or nanocatalysts, they have shown promise as individual fuels in biodiesel-diesel blends, traditional diesel, and biodiesel. Various types of nanoadditives, primarily metals and metal oxides such as cerium (CeO_2), aluminum (Al_2O_3), zinc (ZnO), and titanium (TiO), as well as carbon nanotubes (CNTs), have demonstrated notable effects on exhaust emissions and engine efficiency, with ongoing research continuing to uncover new possibilities [12].

Nanoparticles as Fuel Additives

Because of their substantial energy capacity and large surface-area-to-volume ratio, nanoparticles are considered effective fuel additives [13]. These nanoadditives alter both the physical properties and chemical characteristics of the fuel, including flash points, heating values, viscosity, density, and cetane numbers. The resulting enhancements are evident in shorter ignition delays, higher brake thermal efficiency (BTE), improved fuel consumption rates, increased power output under higher loads, and a reduction in emissions, notably CO, NOx, and HC [14]. Although the use of calcium oxide nanoparticles as a diesel additive has been underexplored, the success of nanoadditives and butanol in emission reduction and performance improvement highlights the relevance of this study. This research aims to probe the combined results of dual fuel additives in varying ratios on performance and emissions. Additionally, fuel blend characterization was conducted according to American society for testing and materials (ASTM) standards, and machine learning techniques were used to model the experimental data to identify the optimal fuel blend for performance and emissions, potentially reducing the costs associated with extensive nano-additive testing.

Emergence of Machine Learning

Researchers are investigating techniques to reduce the number of experimental trials because of the expenses and time commitments involved in experimental engineering research [15]. Although experimental and numerical testing offer important information about how biodiesel fuel mixes affect engine performance and emissions, these techniques frequently require a significant investment in time and money. On the other hand, modeling simulations offer a more efficient method of comprehending fuel blend performance factors, their impact on engine maneuvers, and the exhaust gases released. Machine learning based modeling approaches are the most popular among modeling approaches for simulations.

Building computational models with data access and self-learning capabilities is the key objective of machine learning (ML), which is tasked with determining the best algorithm that can derive solutions from a given response [16]. ML represents an artificial intelligence (AI) application that may effectively and probabilistically develop prognostic and interpolative models for control parameters through self-learning and experience-driven advances without the need for clear programming [17]. AI-powered engine performance and emission modeling provides a flexible model structure that is naturally robust to data with a high degree of uncertainty, making it a useful tool for forecasting and improving engine emissions and performance [18].

Machine learning techniques have been employed to examine the emissions and engine performance of various fuels. When it came to engine response to various hydrogen gas in dual fuel systems, Yıldırım et al. [19] compared the prediction accuracy of artificial neural networks (ANN) to support vector regression (SVR) approaches. They discovered that the ANN outperformed SVR in terms of the total mean absolute percentage error (MAPE) by 53.70% on average.

To create a quick and precise model for forecasting the commencement of ignition in methane-fueled homogenous charge compression ignition engines, Namar et al. [20] used 11 regression models based on ML. Their findings indicated that models built with ML had excellent speed of response, implying that they are appropriate for real-time applications in engines involving electronic parts. Dhahad et al. [21] suggested an AI-driven system for predicting the engine performance, exhaust emissions, and combustion parameters of diesel engines in response to the immediate need to reduce the negative effects of diesel engine emissions on both the environment and human health.

Artificial neural networks have been used to model the effect of biodiesel and diesel mixtures on engine emissions and performance [1]. Their high regression scores (R2), which varied from 0.8663 to 0.9858, showed how well the ANNs predicted a variety of performance and emission metrics. Additionally, when compared with the experimental data, the greatest mean relative error remained below 10%. The ability of ML methods to forecast mining truck fuel consumption was demonstrated by Wang and Zhao [22]. They stated that because of its remarkable prediction accuracy and excellent computing efficiency, extreme gradient boosting (XGBoost) is appropriate for predicting the fuel intake in the mining industry.

Conclusion and Objectives of the Present Study

Numerous studies have investigated how different biodiesel blends affect emissions and engine performance using ML techniques. However, it appears that few studies have been conducted using ML techniques such as XGBoost, decision tree or random forest, and linear regression, and have compared them to specifically analyze how different biodiesel mixes in pure diesel impact engine performance and emissions. The goal of the current work is to deploy ML techniques to estimate and forecast the emissions and performance of a diesel engine. While maintaining steady biodiesel blends, we focused on evaluating various biodiesel combinations. Nevertheless, there are few comparative studies on various ML approaches, such as decision tree or random forest, linear regression, and XGBoost, for engine performance and emissions when diesel and biodiesel mixtures in various proportions are included. By evaluating how well various ML techniques predict engine performance and emissions under various fuel mixtures, we aim to close the research gap.

DATA AND VARIABLES

Preparation of Biodiesel Blends

Waste cooking oil (WCO) is a by-product of the cooking industry and poses environmental hazards when untreated, leading to soil and water pollution. Transesterification processes can produce superior biodiesel with a superior cetane number and an anti-knocking tendency. The biodiesel was blended with diesel in various quantities to test its performance. The results in Table 1 show that WCO biodiesel is comparable to other biofuel families and nearly equals the energy content and density of the base diesel.

Table 1. Compilation of useful properties of different biodiesel blends.

Property	Diesel	B10	B20	B30
Density in g/cc	0.831	0.843	0.852	0.862
Kinematic viscosity in cSt	3.12	3.45	4.87	5.05
Flash point in °C	51	104	106	110
Fire point in °C	57	114	118	121
Cloud point in °C	-10	-3	-3	-2
Calorific value kJ/g	44800	42434	42012	41534

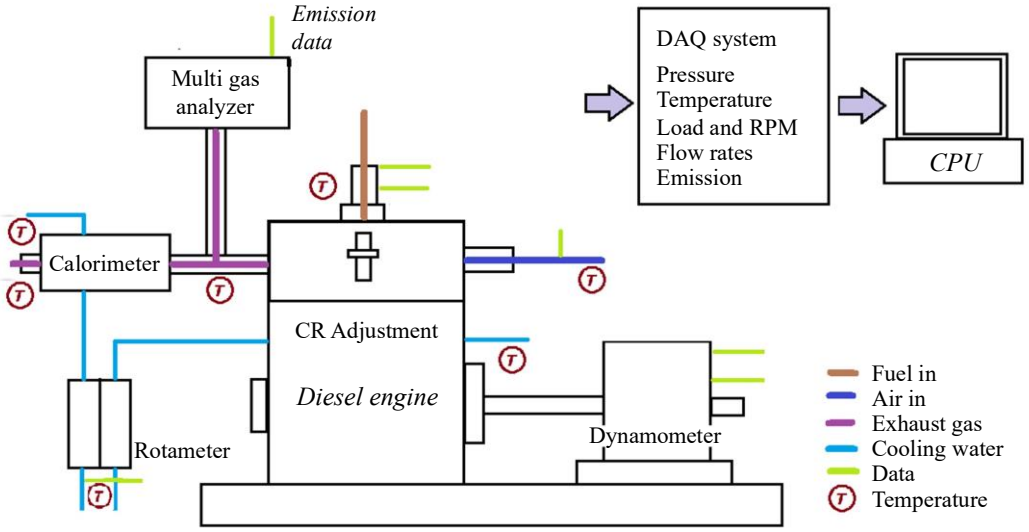

Figure 2. Schematic of test setup and system for data acquisition.

Experimental Setup and Accessories for Engine Testing

The setup illustrated in Figure 2 includes a compression ignition engine with a four-stroke cycle and direct injection into a single cylinder connected to a dynamometer. Instruments measure the flow rates, temperatures, crank angle, cylinder pressure, and load, which are transmitted to a processor via a high-speed data acquisition device. The computer adjusts engine loading using a dynamometer. The setup also used a DIGAS 444N model, a multi-gas analyzer, and a 437°C smoke meter to measure emissions. Table 2 summarizes the configuration of the setup.

Equation (1) was used to determine the brake power specific fuel consumption, brake-specific fuel consumption (BSFC), and brake power-based thermal efficiency, and BTE was calculated using Equation (2), as shown in a previous study [23].

$$BSFC = \frac{1000\,\dot{m}_f}{P_e} \tag{1}$$

Where, the terms are brake-specific fuel consumption (BSFC, g/kWh), hourly fuel consumption (m_f, kg/h), and effective brake power (P_e, kW).

$$BTE = \frac{3600\,P_e}{\dot{m}_f\,LHV} \tag{2}$$

Where, the terms are the thermal efficiency (BTE, %) and the net heating value (LHV, kJ/kg) of the tested fuel. The exhaust gas temperature (EGT) is the third performance parameter in the study of diesel engines, which is an indicator of heat rejected to the surroundings, and it affects pollutant formation. The emission parameters measured were carbon monoxide (CO), carbon dioxide (CO_2), unburned hydrocarbon (HC), oxides of nitrogen (NOx), and smoke intensity (smoke).

Methodology and Model Specifications

This study thoroughly examined the four algorithms used in ML: Decision tree, random forest, linear regression, and XGBoost. All algorithms were run using Google Colab. These ML algorithms were used to predict the engine response variables. There are eight responses that include performance variables, viz. BSFC, BTE, EGT, and emission variables, such as unscorched hydrocarbon (HC), carbon monoxide (CO), nitrogen oxides (NOx), carbon dioxide (CO_2) emissions, and smoke intensity. The training process involved three input variables: the engine load (0%, 25%, 50%, 75%, and 100%), additive concentration (9.50 One hundred PPM), and biodiesel blend ratios (0 and 20%). The study employed a dataset containing 11 clusters of 20 data points (8 × 5 pixels), which were divided into 70% for training and 30% for testing. The following paragraphs briefly describe the three ML algorithms deployed in this study.

Linear Regression Algorithm

A linear regression model was used to predict engine performance and emissions by creating correlations between the input characteristics and target variables. It can analyze past data to predict these correlations, making it suitable for situations with clear links between the features and objectives. Linear regression is computationally efficient owing to its closed-form solution, which enables quick training and prediction. It is suitable for small to moderate datasets and is often preferred over more sophisticated models. However, it is insufficient for capturing complex interactions in engine performance and is sensitive to outliers, thus impacting the model accuracy. This may not be effective for data with nonlinear patterns or complicated feature relationships.

Decision Tree Algorithm

A decision tree regressor is a useful tool for forecasting engine performance and emissions by partitioning data based on input variables, such as engine load, speed, fuel type, and ambient conditions. It learns complex nonlinear relationships between engine features and target variables, offering advantages in engine emission and performance analyses. Decision Tree are ML models that are easier to understand because of their "if-then" rules. They can capture the complex nonlinear interactions between features, making them valuable in engine systems with complex correlations. They do not require feature scaling or normalization, and are less susceptible to outliers than other regression models.

The efficiency of a decision tree regressor is determined by its settings, including max_depth, min_samples_split, min_samples_leaf, max_features, max_leaf_nodes, and ccp_alpha. These parameters help minimize overfitting, optimize splits, and improve model generalization. The max_leaf_nodes parameter reduces the number of leaf nodes, whereas the ccp_alpha parameter reduces tree depth and improves model generalization. These parameters help assess split quality and effectively prune the trees.

Random Forest Algorithm

The random forest regressor is an ensemble learning method used to predict engine performance and emissions and detect intricate, nonlinear associations in data. It is useful for large and complex tasks, such as engine performance forecasts, owing to its effective detection.

Table 2. Qualifications of the test engine.

S.N.	Descriptions	Details
1	Manufacturer	Kirloskar, diesel engine TV-1
2	Diameter and stroke	87.5 mm × 110 mm
3	Rated speed and rated power	1500 rpm and 3.6 kW
4	Compression ratio	17.5:1
5	Displacement volume	661 cc
6	Engine classification	Axis—vertical; number of cylinders—1; number of strokes—4; Cooling medium—water; Injection mode—direct; test speed—constant.

Random Forests are ideal for handling high-dimensional and noisy data by averaging multiple Decision Trees. They provide significant scores for engine settings that affect performance and emissions, aiding engineers in making predictions. Individual trees were less vulnerable to outliers, resulting in more stable models.

XGBoost Algorithm

The XGBoost method is a ML approach that uses the gradient boosting decision tree (GBDT) algorithm for supervised tasks to enhance prediction accuracy. Parallel but distributed processing allows flexible model research. XGBoost combines multiple secondary predictors, such as decision trees, to reduce the residual errors. The final prediction was the sum of the residual forecasts from each tree. The XGBTree approach was used to run all XGBoost models in this study. Certain hyperparameters are used to refine the XGBoost model. max_depth represents the greatest depth of a given tree, which is used to prevent overfitting. max_leaf_nodes had the highest number of end nodes or leaves in a tree. The minimum weighted total of samples necessary for a child is min_child_weight. Similar to the learning rate in the Gradient Boost, the eta parameter value was set between 0.01 and 0.2. The gamma value is tuned to define the minimal drop in loss necessary for a split [24].

Model Evaluation Metrics

The algorithms were evaluated using regression coefficient R2, mean square error (MSE), root mean square error (RMSE), MAPE, and mean absolute error (MAE). These metrics help determine the predicted error rates and model performance based on the average absolute variance of the dataset.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y - \hat{y}| \tag{3}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2 \tag{4}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2} \tag{5}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2} \tag{6}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}}{y_i}\right| \tag{7}$$

The mean squared error (MSE) and root mean squared error (RMSE) are metrics used to measure model accuracy and performance. The MSE measures the difference between the initial and expected outcomes, whereas the RMSE is the error rate. The R2 values indicate robust predictive associations. RMSE and MAE are errors, with lower values indicating a better model performance.

RESULTS AND DISCUSSION

Evaluation of Prediction of Engine Performance

Figure 3 shows a line plot comparing the predictive performance of the four different regression models (decision tree, random forest, linear regression, and XGBoost) across a range of values for the target features. The y-axis, labeled “y_pred,” represents the predicted output of each model, which represents the engine performance values.

Evaluation of Prediction of Engine Emission Performance

Figures 4 and 5 depict line plots comparing the predictive performance of the four different regression models (decision tree, random forest, linear regression, and XGBoost) across an assortment of target features. The y-axis, labeled “y_pred,” shows the predicted outputs of each model, which are the corresponding engine emission values.

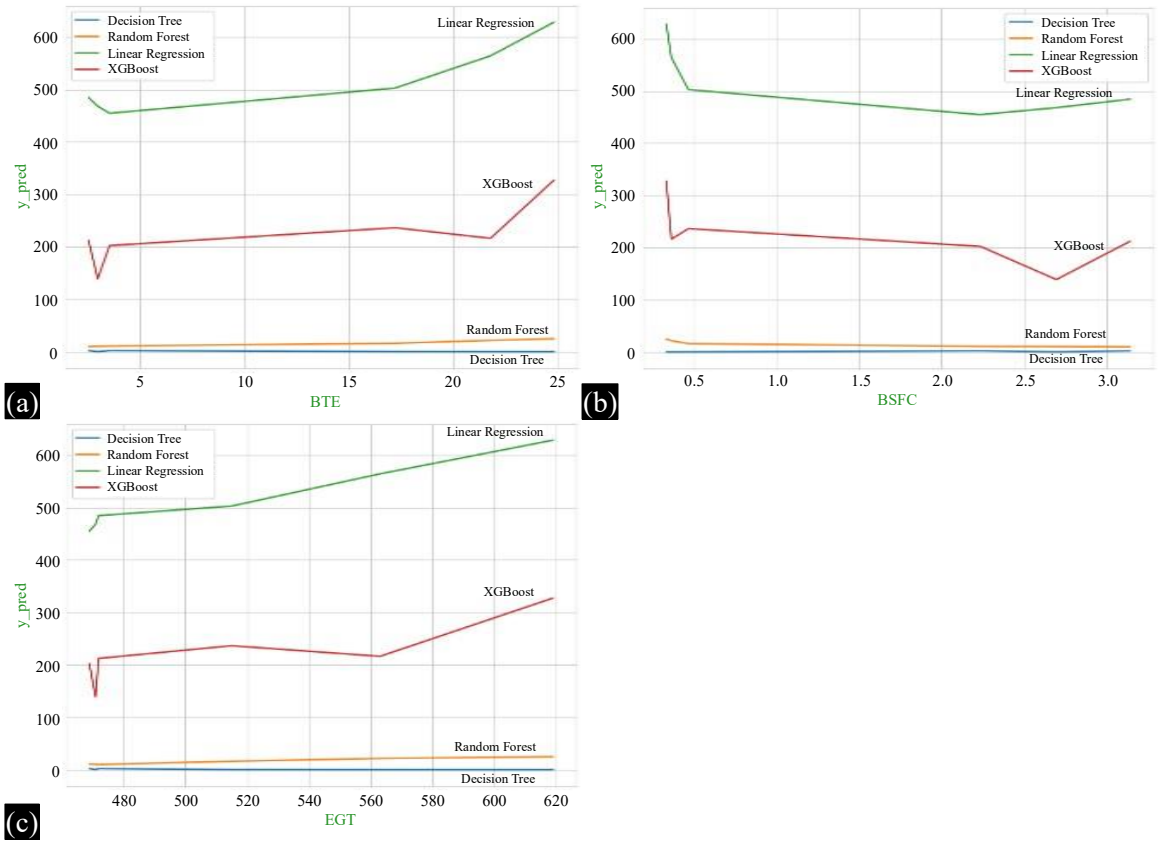

Figure 3. Engine performance (BTE, BSFC and EGT) using different models.

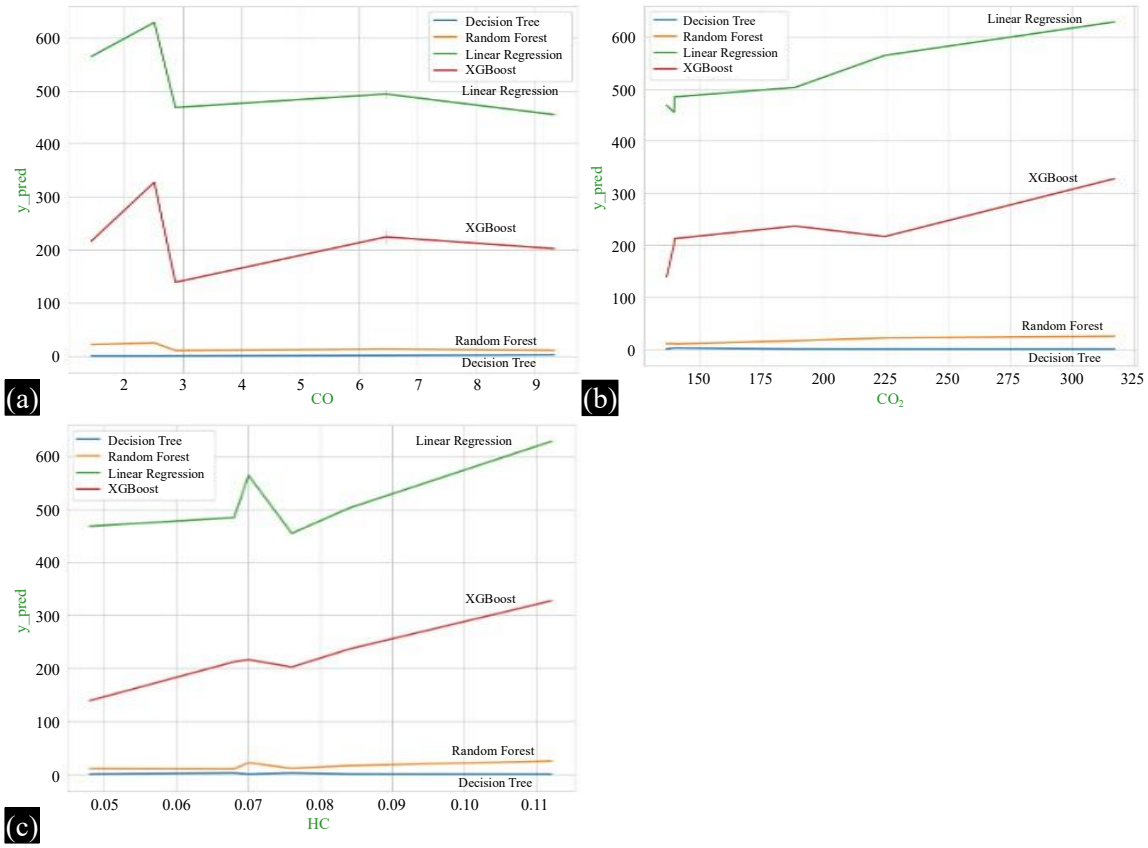

Figure 4. Emission performance (CO, CO_2 and HC) using different models.

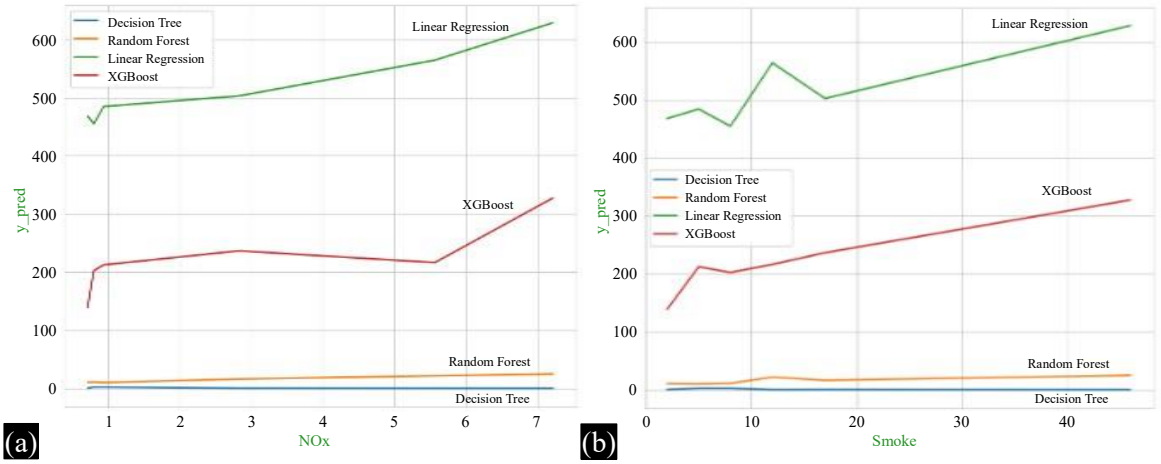

Figure 5. Emission performance (NOx and smoke) using different models.

Observations from Prediction of Engine Performance and Emission Plots

The decision tree (blue) model produces low and relatively stable predictions across the range of all features. This suggests that the decision tree might be underfitting, as it does not show much variation in y_pred in response to the features. The random forest (orange) model also showed a relatively stable trend with predictions near zero across all feature values. Similar to the decision tree, it exhibited little response to changes in features, suggesting a similar trend of underfitting.

Linear regression (green) displays a declining trend in y_pred as features such as BSFC increase, with predictions starting high and gradually decreasing. This response indicates that linear regression attempts to fit a linear relationship between BSFC and the target, but its linear nature may limit its performance on these data. XGBoost (red) shows a more variable trend with some fluctuations, starting with higher y_pred values around 300 and showing some decline, which is followed by an upsurge as the BSFC values increase. This indicates that XGBoost was more responsive to variations in BSFC, likely capturing more complex patterns in the data.

The output shows that different regression models respond differently to variations in the response variable. More complex models, particularly XGBoost, appear to capture more variation in the response variable, potentially indicating a better fit. Meanwhile, the decision tree and random forest models exhibit a lower sensitivity to changes in response variables, which might indicate underfitting. Linear regression captures a monotonic decline in y_pred, probably because of its linear structure.

XGBoost may be more suitable for capturing nonlinear patterns. Decision tree and random forest models might require further tuning (e.g., more trees for random forest) to improve performance. Linear regression may be limited in capturing complex relationships in this dataset, as indicated by its linear trend in y_pred.

Model Evaluation

Table 3 presents the evaluation metrics for the four regression simulations: decision tree, random forest, linear regression, and XGBoost. The key metrics analyzed included the R^2 score, MSE, MAE, MAPE, and RMSE.

R^2 Score

XGBoost achieves the peak R^2 score (0.8451), demonstrating that it explains approximately 84.51% of the variance in the target variable, making it the optimum model for this metric. The random forest and decision tree had lower R^2 scores (0.5610 and 0.4698, respectively), indicating weaker explanatory power. Linear regression achieved a moderate R^2 score of 0.5430, but it was still outperformed by XGBoost.

Table 3. Model evaluation metrics summary.

Evaluation matrix	Models deployed for prediction			
	Decision Tree	***Random Forest***	***Linear regression***	***XGBoost***
R^2 Score	0.4698	0.5610	0.5430	0.8451
MSE	485.2605	527.8846	235.3589	255.7107
MAE	9.8924	11.4854	7.4464	5.6603
MAPE	0.3791	0.4906	0.7519	0.1646
RMSE	12.6964	12.9782	8.9613	7.4904

Mean Squared Error

Linear regression had the lowest MSE (235.3589), suggesting that its average squared error was smaller than that of the other models. This implies that it performs well in terms of the overall magnitude of the prediction error. XGBoost follows closely with an MSE of 255.7107, indicating that it also performs well on this metric. The decision tree and random forest had higher MSE values, with random forest having the highest (527.8846), indicating relatively larger prediction errors.

Mean Absolute Error

XGBoost had the lowest MAE (5.6603), showing that, on average, its absolute prediction error was the smallest, making it the best-performing model in terms of absolute deviation from actual values. Linear regression also performed well with an MAE of 7.4464. The decision tree and random forest show larger errors, with the random forest having the highest MAE (11.4854).

Mean Absolute Percentage Error

XGBoost shows the lowest MAPE (0.1646), indicating a relatively small percentage error on average and suggesting that its forecasts are proportionally nearer to actual values. Decision tree and random forest had MAPE values of 0.3791 and 0.4906, respectively, whereas linear regression had the maximum MAPE (0.7519), indicating less accurate proportional predictions compared to XGBoost.

Root Mean Squared Error

XGBoost had the lowest RMSE (7.4904), suggesting better prediction accuracy and stability in capturing large errors than the other models. Linear regression had a reasonably low RMSE (8.9613), but it was outperformed by XGBoost. The decision tree and random forest had higher RMSE values (12.6964 and 12.9782, respectively), indicating a relatively poor performance.

In summary, XGBoost is the most accurate and robust model for forecasting engine performance and emissions in a dataset, outperforming other models such as Decision Trees and Random Forests. XGBoost provided a better balance across all metrics, making it the recommended model for practical applications. Linear Regression has the lowest MSE, but Decision Trees and Random Forests show lower performance, suggesting that they may be underfitting or less capable of capturing data complexity.

CONCLUSION AND FUTURE SCOPE

Conclusion

This study investigated the exhaust emissions and performance of a diesel engine blended with biodiesel and nanoparticle additives. Various modeling techniques were used, with input parameters including fuel type, nanoparticle concentration, and engine load and response variables including BTE, BSFC, EGT, CO, CO_2, NOx, HC, and smoke intensity. The results of this study are summarized as follows:

1. Different regression models respond differently to variations in response variables.
2. Complex models like XGBoost capture more variation, suggesting better fit.
3. Decision tree and random forest models show lower sensitivity, indicating underfitting.
4. Linear regression captures monotonic decline, XGBoost captures nonlinear patterns.
5. XGBoost model is most accurate and robust for forecasting engine performance and emissions.

6. XGBoost provides better balance across all metrics, making it recommended for practical application.

Scope for Future Work

Further tuning of the decision tree and random forest models may improve the performance. Advanced ensemble models can be used in the future to modeling the emission behavior and engine performance of diesel engines.

Declaration of Interest

The authors declare no conflicts of interest regarding the publication of this manuscript.

Acknowledgement

NA.

REFERENCES

1. Aydın M, Uslu S, Bahattin Çelik MB. Performance and emission prediction of a compression ignition engine fueled with biodiesel-diesel blends: A combined application of ANN and RSM based optimization. Fuel. 2020;269:117472. doi:10.1016/j.fuel.2020.117472.
2. Ağbulut Ü, Gürel AE, Sarıdemir S. Experimental investigation and prediction of performance and emission responses of a CI engine fuelled with different metal-oxide based nanoparticles–diesel blends using different machine learning algorithms. Energy. 2021;215:119076. doi:10.1016/j. energy.2020.119076.
3. Uslu S, Celik MB. Combustion and emission characteristics of isoamyl alcohol-gasoline blends in spark ignition engine. Fuel. 2020;262:116496. doi:10.1016/j.fuel.2019.116496.
4. Salvi BL, Subramanian KA, Panwar NL. Alternative fuels for transportation vehicles: A technical review. Renew Sustain Energy Rev. 2013;25:404–19. doi:10.1016/j.rser.2013.04.017.
5. Karikalan L, Jacob S, Baskar S, Venugopal S. Analyzing the influence of varied fuel injection pressure on diesel engine fueled with Karanja biodiesel. Mater Today Proc. 2020;21:291–4. doi:10.1016/j.matpr.2019.05.439.
6. Simsek S. Effects of biodiesel obtained from Canola, safflower oils and waste oils on the engine performance and exhaust emissions. Fuel. 2020;265:117026. doi:10.1016/j.fuel.2020.117026.
7. Elkelawy M, Alm-Eldin Bastawissi H, Esmaeil KK, Radwan AM, Panchal H, Sadasivuni KK, et al. Experimental studies on the biodiesel production parameters optimization of sunflower and soybean oil mixture and DI engine combustion, performance, and emission analysis fueled with diesel/biodiesel blends. Fuel. 2019;255:115791. doi:10.1016/j.fuel.2019.115791.
8. Gozmen Şanli B, Uludamar E, Özcanli M. Evaluation of energetic-exergetic and sustainability parameters of biodiesel fuels produced from palm oil and opium poppy oil as alternative fuels in diesel engines. Fuel. 2019;258:116116. doi:10.1016/j.fuel.2019.116116.
9. Holechek JL, Geli HME, Sawalhah MN, Valdez R. A global assessment: Can renewable energy replace fossil fuels by 2050? Sustainability. 2022;14:4792. doi:10.3390/su14084792.
10. Hoang AT, Nižetić S, Olcer AI, Ong HC, Chen WH, Chong CT, et al. Impacts of COVID-19 pandemic on the global energy system and the shift progress to renewable energy: Opportunities, challenges, and policy implications. Energy Policy. 2021;154:112322. doi:10.1016/j.enpol.2021. 112322. PMID: 34566236.
11. Ahmad S, Jafry AT, Haq M, Abbas N, Ajab H, Hussain A, et al. Performance and emission characteristics of second-generation biodiesel with oxygenated additives. Energies. 2023;16. doi:10.3390/en16135153.
12. Ağbulut Ü, Karagöz M, Sarıdemir S, Öztürk A. Impact of various metal-oxide based nanoparticles and biodiesel blends on the combustion, performance, emission, vibration and noise characteristics of a CI engine. Fuel. 2020;270:117521. doi:10.1016/j.fuel.2020.117521.
13. Altammar KA. A review on nanoparticles: Characteristics, synthesis, applications, and challenges. Front Microbiol. 2023;14:1155622. doi:10.3389/fmicb.2023.1155622. PMID: 37180257.

14. Abdallah AM, Abdel-Rahman AA, Elwardany AE. Analysis of the impact of different nanoparticle metal oxides as fuel additives in compression ignition engine performance. Energy Rep. 2020;6:99–105. doi:10.1016/j.egyr.2020.12.016.
15. Montgomery DC. Design and Analysis of Experiments. New York: John Wiley & Sons; 2017.
16. Pattnaik P, Sharma A, Choudhary M, Singh V, Agarwal P, Kukshal V. Role of machine learning in the field of fiber reinforced polymer composites: A preliminary discussion. Mater Today Proc. 2021;44:4703–8. doi:10.1016/j.matpr.2020.11.026.
17. El-Seesy AI, Waly MS, He Z, El-Batsh HM, Nasser A, El-Zoheiry RM. Influence of quaternary combinations of biodiesel/methanol/n-octanol/diethyl ether from waste cooking oil on combustion, emission, and stability aspects of a diesel engine. Energy Convers Manag. 2021;240:114268. doi:10.1016/j.enconman.2021.114268.
18. Roy S, Banerjee R, Bose PK. Performance and exhaust emissions prediction of a CRDI assisted single cylinder diesel engine coupled with EGR using artificial neural network. Appl Energy. 2014;119:330–40. doi:10.1016/j.apenergy.2014.01.044.
19. Yıldırım S, Tosun E, Çalık A, Uluocak İ, Avşar E. Artificial intelligence techniques for the vibration, noise, and emission characteristics of a hydrogen-enriched diesel engine. Energy Sources Part A. 2019;41:2194–206. doi:10.1080/15567036.2018.1550540.
20. Namar MM, Jahanian O, Koten H. The start of combustion prediction for methane-fueled HCCI engines: Traditional vs. machine learning methods. Math Probl Eng. 2022;2022:1–11. doi:10.1155/2022/4589160.
21. Dhahad HA, Hasan AM, Chaichan MT, Kazem HA. Prognostic of diesel engine emissions and performance based on an intelligent technique for nanoparticle additives. Energy. 2022;238:121855. doi:10.1016/j.energy.2021.121855.
22. Wang H, Zhao W. ARIMA model estimated by particle swarm optimization algorithm for consumer price index forecasting. In: Deng H, Wang L, Wang FL, Lei J, editors. Artif Intell Comput Intell. Berlin: Springer; 2009. p. 48–58. doi:10.1007/978-3-642-05253-8_6.
23. Agrawal T, Gautam R, Agrawal S, Singh V, Kumar M, Kumar S. Optimization of engine performance parameters and exhaust emissions in compression ignition engine fueled with biodiesel-alcohol blends using Taguchi method, multiple regression and artificial neural network. Sustain Futures. 2020;2:100039. doi:10.1016/j.sftr.2020.100039.
24. Zivkovic M, Jovanovic L, Ivanovic M, Bacanin N, Strumberger I, Joseph PM. XGBoost hyperparameters tuning by fitness-dependent optimizer for network intrusion detection. In: Sharma H, Shrivastava V, Kumari Bharti K, Wang L, editors. Communication and Intelligent Systems. Singapore: Springer; 2022. (Lecture Notes in Networks and Systems; vol. 461). doi:10.1007/978-981-19-2130-8_74.

