Contents list available at CBIORE journal website

International Journal of Renewable Energy Development

Journal homepage: https://ijred.cbiore.id

Research Article

Application of supervised machine learning and Taylor diagrams for prognostic analysis of performance and emission characteristics of biogas-powered dual-fuel diesel engine

Khac Binh Le[a], Minh Thai Duong[b], Dao Nam Cao[b], Van Vang Le[c,*]

[a]*Vinh University of Technology Education, 117 Nguyen Viet Xuan Street, Hung Dung Ward, Vinh City, Viet Nam.*

[b]*Institute of Mechanical Engineering, Ho Chi Minh City University of Transport, Ho Chi Minh City, Viet Nam.*

[c]*Institute of Maritime, Ho Chi Minh City University of Transport, Ho Chi Minh City, Viet Nam.*

Abstract. In the ongoing search for an alternative fuel for diesel engines, biogas is an attractive option. Biogas can be used in dual-fuel mode with diesel as pilot fuel. This work investigates the modeling of injecting strategies for a waste-derived biogas-powered dual-fuel engine. Engine performance and emissions were projected using supervised machine learning methods including random forest, lasso regression, and support vector machines (SVM). Mean Squared Error (MSE), R-squared (R^2), and Mean Absolute Percentage Error (MAPE) were among the criteria used in evaluations of the models. Random Forest has shown better performance for Brake Thermal Efficiency (BTE) with a test R^2 of 0.9938 and a low test MAPE of 3.0741%. Random Forest once more exceeded other models with a test R^2 of 0.9715 and a test MAPE of 4.2242% in estimating Brake Specific Energy Consumption (BSEC). With a test R^2 of 0.9821 and a test MAPE of 2.5801% Random Forest emerged as the most accurate model according to carbon dioxide (CO_2) emission modeling. Analogous results for the carbon monoxide (CO) prediction model based on Random Forest obtained a test R^2 of 0.8339 with a test MAPE of 3.6099%. Random Forest outperformed Linear Regression with a test R^2 of 0.9756% and a test MAPE of 7.2056% in the case of nitrogen oxide (NOx) emissions. Random Forest showed the most constant performance overall criteria. This paper emphasizes how well machine learning models especially Random Forest can prognosticate the performance of biogas dual-fuel engines.

Keywords: Biogas; Alternative fuel; Supervised machine learning; Lasso regression; Random Forest; Taylor diagram

Received: 25th June 2024; Revised: 19th October 2024; Accepted: 27th October 2024; Available online: 15th November 2024

1. Introduction

SDG 7 seeks to guarantee that everyone can afford modern, reasonably priced, environmentally friendly energy, these targets can be achieved through various sustainable practices (Nguyen *et al.*, 2024; "Tracking SDG 7 – The Energy Progress Report 2022," n.d.). The use of biogas is one such option. Biogas is generated from organic sources like food waste, sewage, and agricultural trash (Hoang *et al.*, 2022; Naghavi *et al.*, 2020; The-Thanh *et al.*, 2019). That is the reason it is termed as a sustainable substitute for fossil fuels. By lowering reliance on conventional fuels such as charcoal or firewood, the use of biogas for cooking, heating, and electricity all help to increase energy availability (Fransiscus and Simangunsong, 2021; Hidayanti *et al.*, 2021; Sharma *et al.*, 2023b). Other SDGs including Zero Hunger (SDG 2), which lowers food waste and supplies fertilizer, and Climate Action (SDG 13), which lowers greenhouse gas emissions, benefit from biogas generating as well. By offering a dependable and sustainable energy source, boosting energy access, and so supporting more general sustainability goals, biogas helps to meet SDG 7 (Lohani *et al.*, 2021; Rocha-Meneses *et al.*, 2023; Runyowa and Fourie, 2021).

With good fuel economy, durability, and great torque, diesel engines run generators, heavy equipment, and vehicles (Cunanan *et al.*, 2021; Hoang, 2018). Modern diesel engines use technologies including direct injection and turbocharging for lower emissions and improved performance (Changxiong *et al.*, 2023; Hoang, 2021). They do, however, spew more nitrogen oxides and particulate matter, which aggravates air pollution. By burning more cleanly and effectively, alternative fuels such as biodiesel, alcohol, natural gas, and hydrogen can help to lower emissions (Abdullah *et al.*, 2019; Hoang *et al.*, 2023; Ramalingam *et al.*, 2023; Serbin *et al.*, 2022; Veza *et al.*, 2022b; Zhao *et al.*, 2020). They also lessen engine part deterioration and increase engine lifetime. Alternative fuels thus help to create a less polluting and more sustainable transportation network (E *et al.*, 2017; Rahman *et al.*, 2023; Sakthivel *et al.*, 2018).

There are two ways to use biogas in an internal combustion engine. Either as single fuel model in which biogas is used with spark ignition or in dual-fuel mode (DFM) in which biogas remains the main fuel while a tiny quantity of liquid fuel (diesel/biodiesel) is employed as pilot fuel to initiate the combustion (Feroskhan *et al.*, 2018; Goyal *et al.*, 2023; Jamei *et al.*, 2024; Nguyen-Thi and Bui, 2023). Dual-fuel engines emit less harmful pollutants than conventional diesel engines, so

* Corresponding author:
Email: levanvang@ut.edu.vn (V.V. Le)

https://doi.org/10.61435/ijred.2024.60724
ISSN: 2252-4940/© 2024.The Author(s). Published by CBIORE

benefiting the environment (Boretti, 2019; Bui *et al*., 2024). This is because the lower carbon content of biogas consumed in DFM causes lower emissions of glasshouse gases and other pollutants (Mahla *et al*., 2018; Sharma *et al*., 2023c). Combining the main and secondary fuels also increases the efficiency of burning the main fuel, thus saving money and lessening pollution (Bui *et al*., 2021; Meng *et al*., 2020). Using biogas as a fuel has one of the primary environmental advantages as well. Unlike fossil fuels, biogas produces no harmful glasshouse gases like methane or carbon dioxide, both of which greatly contribute to global warming (Bui *et al*., 2023; Czekała, 2022). Using biogas as a fuel in dual-fuel engines can therefore significantly reduce carbon emissions and help to slow down the effects of climate change (Awe *et al*., 2017; Mao *et al*., 2015; Parsaee *et al*., 2019). The low cost of dual-fuel engines driven by biogas is another noteworthy advantage. Biogas is easily available and reasonably priced since it is a by-product of organic waste (Bui *et al*., 2022; Kapoor *et al*., 2020). Consequently, running engines on biogas could drastically reduce their fuel cost. Moreover, employing less fuel to reach the same degree of performance, dual-fuel engines also have greater fuel economy than conventional engines (Feroskhan and Ismail, 2017; Vasan *et al*., 2024).

The low cost and environmental friendliness of biogas-powered dual-fuel engines have helped them to become rather popular recently. Still, these engines offer several challenges and problems. Regarding biogas quality and availability, running dual-fuel engines powered by biogas creates major challenges (Ardolino *et al*., 2021). The type of feedstock and the source of the gas will affect the quality of biogas quite significantly (Karne *et al*., 2023). Biogas contaminants including hydrogen sulphide could cause engine parts to fail. Biogas burns more slowly than diesel since its ignition delay is significant. All of which can affect engine efficiency are knock, misfiring, and lower power output resulting from this (Khayum *et al*., 2020). Because biogas has a lower energy density than diesel, the engine needs more of it to generate the same power level (Prabhu *et al*., 2021). This could lower the range and increase fuel consumption, thus influencing the general engine performance. The working of biogas-powered engines in dual fuel mode is a highly complex phenomenon (Ahmad *et al*., 2024a). Modeling in such circumstances becomes a difficult and manually intensive process (Ahmad *et al*., 2024b). Machine learning (ML) techniques are attractive options and are being explored by researchers for such complex engineering problems. Several investigators explored the applications of ML in model prediction of dual-fuel engines (Sharma *et al*., 2023a). For challenging engineering problems, modern ML techniques including Lasso regression, Support Vector Machines (SVM), and Random Forests (RF) offer great benefits to several scientific and technological fields such as energy, fuels, education, transportation, society, and healthcare (Haque *et al*., 2024; Nisa *et al*., 2023; Puri *et al*., 2023; P. ; Sharma *et al*., 2022; P. Sharma *et al*., 2022; Sharmila *et al*., 2024). These methods are progressively preferred for several reasons. For regularisation and feature selection especially lasso regression is quite helpful. Lasso regression reduces less important coefficients to zero, so helping to identify the most relevant elements in complex systems where many input variables can affect results. This simplifies the model and increases interpretability without sacrificing predictive accuracy (García-Nieto *et al*., n.d.; Mohammad *et al*., 2022). This method guarantees that only the most important variables are kept for researchers handling high-dimensional data, producing more concentrated and effective models. Particularly in cases involving non-linear relationships, SVM is fit for classification tasks (Meenal and Selvakumar, 2018). Since SVMs can manage high-dimensional spaces, they are efficient in situations when the input variables are many and their interactions are complicated. SVMs provide strong forecasts even in small datasets by optimizing the margin between data classes. Optimizing engine performance depends on this quality since the exact classification of operational states can greatly affect efficiency and emissions (Hao *et al*., 2020; Rao *et al*., 2022).

The research on biogas-powered dual-fuel engines reveals notable advancements in alternative fuel use to lower emissions and improve engine performance. Often with an eye toward combustion optimization and emission reduction, several studies have examined the performance and emission characteristics of biogas in dual-fuel engines (Bora and Saha, 2016; Sahoo *et al*., 2009; Tira *et al*., 2014; Verma *et al*., 2017). Machine learning (ML) techniques have also been applied to simulate intricate engine operations. Still, there are gaps. Most of the current research depends on a few ML techniques, such as simple regression models or neural networks (Bhatt and Shrivastava, 2021; Bietresato *et al*., 2015; Onukwuli *et al*., 2021; Patnaik *et al*., 2024; Usman *et al*., 2021; Veza *et al*., 2022a). Few studies have methodically evaluated engine performance and emissions with sophisticated supervised learning models like Lasso Regression, SVM, and Random Forest. Furthermore, little study has been done using these models for injection strategy modeling, which is very essential for dual-fuel engine optimization. This work aims to close these gaps by using Lasso Regression, SVM, and Random Forest to simulate injection methods for a dual-fuel engine running biogas. The work seeks to maximize engine performance, lower emissions, and find the best ML method. The three ML models will be applied, a comprehensive dataset from engine trials will be created, and their performance will be assessed using MSE, R^2, and MAPE. This method will provide a more thorough understanding of how well certain ML methods optimize engines.

2. Materials and methods

2.1 Test setup and test fuel

In the present study, a diesel engine fitted with a variable compression arrangement was employed. The test engine was single cylinder 3.5 kW rated power at 1500 rpm. It was a water-cooled engine. The engine has arrangements for fuel flow measurement, cooling water, airflow, biogas flow, and engine speed. A gas mixer was used for mixing the incoming biogas with air before it entered the engine cylinder. In this study, a 6 m3 fixed dome-type biogas generation was used. The feedstock for biogas generation was food waste and animal manure. Within the dome, these organic materials underwent anaerobic digestion under which microorganisms broke them down without oxygen. With methane the main component of biogas, this process generated a mix of methane and carbon dioxide. Stored in the dome, the produced biogas was then fed into the engine for fuel. By using animal manure and food waste as feedstock, one found a sustainable solution that lowers waste and generates renewable energy. The diesel engine's dual-fuel running depended critically on the produced biogas. It's mixing with air and later combustion let some diesel fuel be replaced, so lowering emissions and maybe increasing engine efficiency. The gas was transported from the gas generation site to the lab in a special-purpose balloon. The engine test setup and other arrangements are depicted in Figure 1. The properties of biogas and diesel used as pilot fuel are listed in Table 1.

2.2 Test procedure

The engine was initially run on diesel only at a low load for 30 minutes to ensure that the temperature of lubricating oil was stabilized. The engine testing was started at 23° before the top

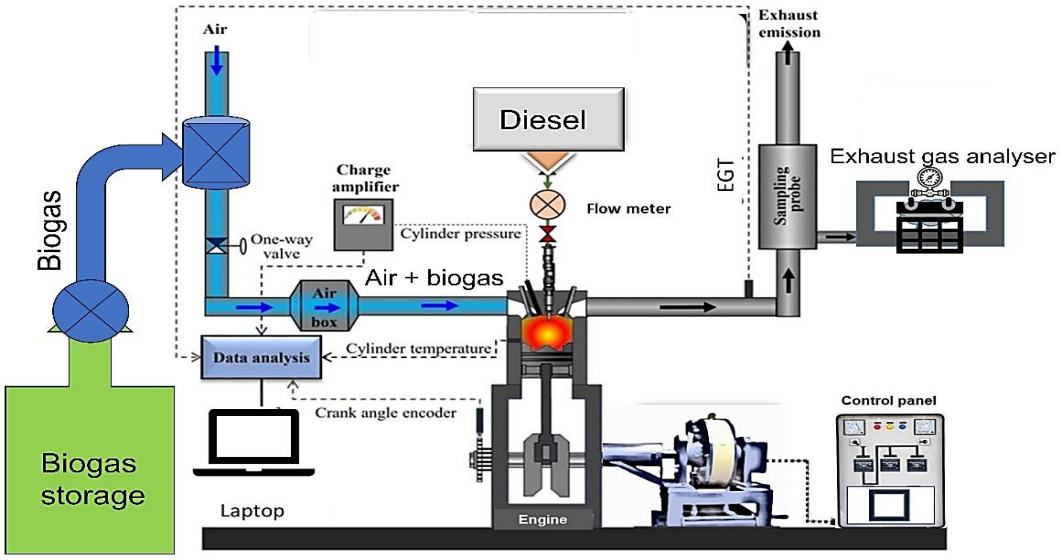

Fig. 1 Test engiune setup

Table 1
Test fuel properties

Property	Specifications	Diesel	Biogas
Lower Heating Value (MJ/kg)	ASTM D240	42.2	21240 ± 100
Cetane Number	ASTM D613	46	--
Flash Point (°C)	ASTM D93	76	--
Density (kg/m^3)	ASTM D4052	843	0.92 ± 0.16
Fire Point (°C)	ASTM D92	85	--
Auto ignition point (°C)		281	812 ± 5

dead center (BTDC) and 17.5 CR to collect baseline data. The engine load was varied from low load to fuel engine load by increasing the engine load by 20% each time. The same approach was followed at different fuel injection timing and different engine loads. The data was collected for emission and other values like engine speed, fuel flow rates, and airflow rates were used for the calculation of engine performance values like BTE, and BSEC, while PCP was directly measured using a piezoelectric pressure sensor installed on the engine head.

3. Soft computing methods

3.1 Lasso regression

Lasso Regression, or Least Absolute Shrinkage and Selection Operator, is a linear regression technique that includes regularization. It appends to the loss function a penalty equal to the absolute value of the coefficient magnitude. This penalty term can shrink some to zero and limit the coefficients, so acting as variable selection (Li *et al.*, 2023). Lasso simplifies the model and helps to prevent overfitting, so it is helpful when there are lots of features. When data shows multicollinearity or when researchers want to find a subset of pertinent predictors, it is especially helpful (Ayyıldız and Murat, 2024). Lasso lowers the variance without an appreciable increase in bias by cutting the number of features, making the model more interpretable. Usually chosen by cross-valuation, the penalty term, which is under control by a tuning parameter determines the degree of the regularization. Python libraries such as sci-kit-learn offer quick Lasso Regression implementations, so enabling researchers to apply this method readily to their data. Lasso Regression is a useful instrument overall for creating parsimonious models balancing interpretability with accuracy (García-Nieto *et al.*, n.d.; Yu *et al.*, 2024).

3.2 Support vector machines

Support Vector Machines (SVM) is an efficient supervised learning algorithm used for both classification and regression tasks. The main objective is to identify the hyperplane that best separates the data points of different classes in the feature space (Najafi *et al.*, 2016). In the case of a regression problem, it is known as Support Vector Regression (SVR). The approach in regression seeks to minimize the error by fitting the best line inside a predefined margin, epsilon. Because of its kernel trick, which lets SVM handle non-linearly separable data by mapping it into higher-dimensional space where a linear separator can be found. SVM is flexible and effective in high-dimensional spaces (Cortes and Vapnik, 1995; Tanveer *et al.*, 2022). Utilizing regularizing parameters, which manage the trade-off between the complexity of the model and its performance on the training data, SVMs offer flexibility and resistance to outliers. SVMs are especially suited for applications when the number of dimensions exceeds the number of samples (Pisner and Schnyer, 2020; Shi *et al.*, 2012).

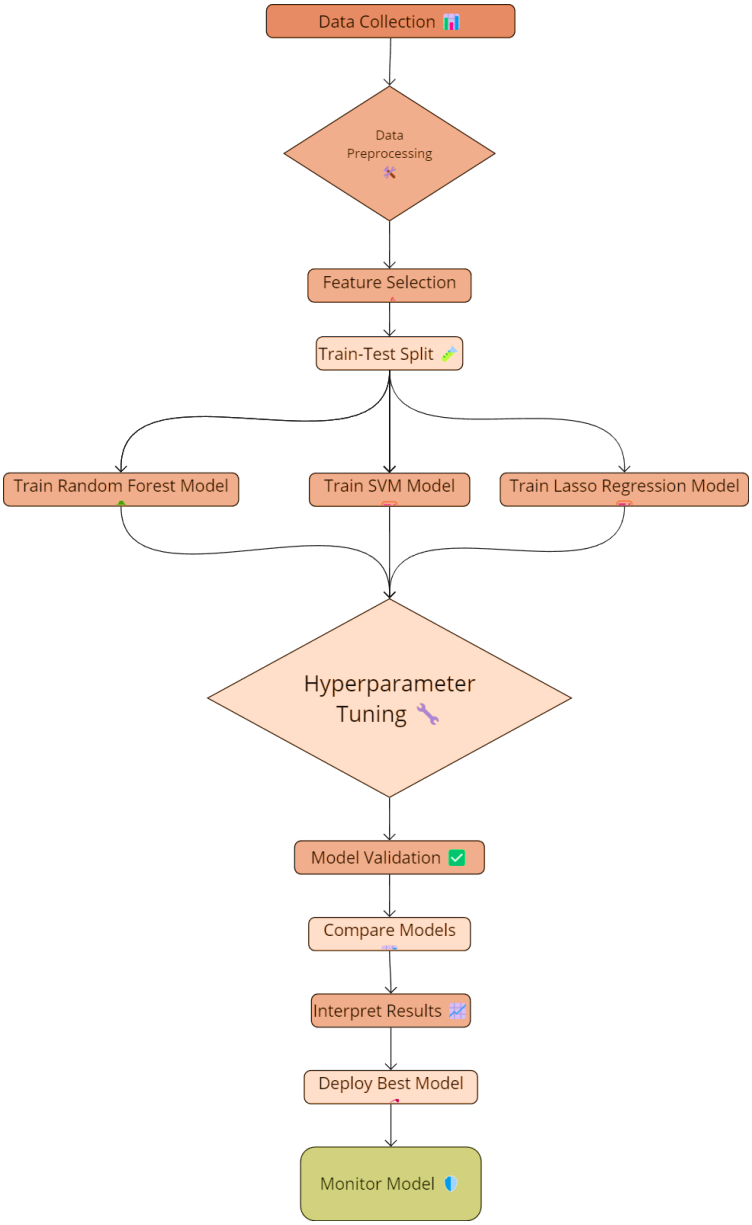

Fig. 2 Flow chart for ML implementation

3.3 Support vector machines

Designed as an ensemble learning method, Random Forest generates several decision trees during training and aggregates them to produce a more reliable prediction. Every tree found in the forest is constructed using a random subset of features and a random subset of the training data (Breiman, 2001). This randomness helps to reduce overfit and increases the generalizing capability of the model. Random Forest uses a majority vote in classification and averages the predictions of individual trees in regression projects. This approach is also defined by high accuracy, dependability, and ability to manage large datasets with more dimensionality (Schonlau and Zou, 2020; Walker *et al.*, 2022). Random forests expose information on feature importance, so guiding feature selection; they are less prone than single decision trees to overfit. In the absence of major parameter tuning, they can also control missing data and maintain performance (Gholizadeh *et al.*, 2020)

3.4 Model evaluation

Once the models are developed, these can be evaluated and compared using statistical indices like coefficient of determination, mean squared error, and mean absolute percentage error. The coefficient of determination, or R^2, measures the share of the dependent variable variation that is predicted from the independent variables. Its value lies between 0 to 1; 0 denotes no forecasting ability and 1 denotes the best possible. Mean Squared Error (MSE) is the computation of the expected and actual values' average squared difference. MSE reveals the degree of match between the projections and the actual data, low values fit more precisely. Mean Absolute Percentage Error (MAPE) is a reported percentage value indicating the degree of prediction accuracy, it finds the average absolute percentage difference between the projected and actual values. Low values point to improved precision. The following expressions were used for calculation (Jamei *et al.*, 2020; Kanti *et al.*, 2023; Kumar Kanti *et al.*, 2023; Sharma and Bora, 2023):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (1)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (2)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \quad (3)$$

Herein, y_i denotes the measured value in the case of ith observation, $\hat{y}_i$ denotes the predicted value for ith observation, n is the total number of observations under considered, $\bar{y}_i$ denotes the mean value of observations.

3.5 Taylor's diagrams for ML model comparison

Taylor diagrams help to evaluate and compare ML models. They present a quick graphical summary. They weigh several aspects of model performance. Among these are correlation; standard deviation; and root-mean-square error (RMSE). Taylor diagrams combine these values on one plot. This makes one more able to assess models. The diagram's angle points to the correlation. Radial distance reveals the standard deviation. Co ntours capture RMSE. At the reference point, one finds a perfect model. Better models are those almost at this point (Elvidge *et al*., 2014; Simão *et al*., 2020). This method underlines differences between models quickly. It is quite helpful for comparing several models. It guides the choice of the model with optimal performance. Difficult performance criteria can be simplified with Taylor diagrams. They present an unambiguous graphic assessment. In machine learning, they are practical tools. Their influence is felt in several spheres. They help to evaluate and compare models (Taylor, 2001).

4. Results and discussion

4.1 Correlational analysis of data

The data gathered through engine testing at different operational settings was analyzed through correlational analysis. The correlation values calculated are listed in Table 2. The correlational heatmap is depicted in Figure 3. An important new understanding of the interplay among fundamental engine parameters is revealed by the correlation analysis of the biogas-powered engine running in dual-fuel mode. A strongly positive connection between the load (%) and Brake Thermal Efficiency (BTE) (0.97) shows that higher engine loads greatly improve efficiency. This link also included carbon dioxide (CO_2) emissions (0.88), wherein greater CO_2 generation was followed by increasing load on the engine. But load revealed a significant negative association with Brake Specific Energy Consumption (BSEC) (-0.89), meaning that larger loads result in lower energy consumption per unit of output, hence improving general engine performance. With NOx emissions (0.92) and CO_2 emissions (0.93), the BTE showed a substantial positive connection. This implies a trade-off between higher emissions, especially NOx,

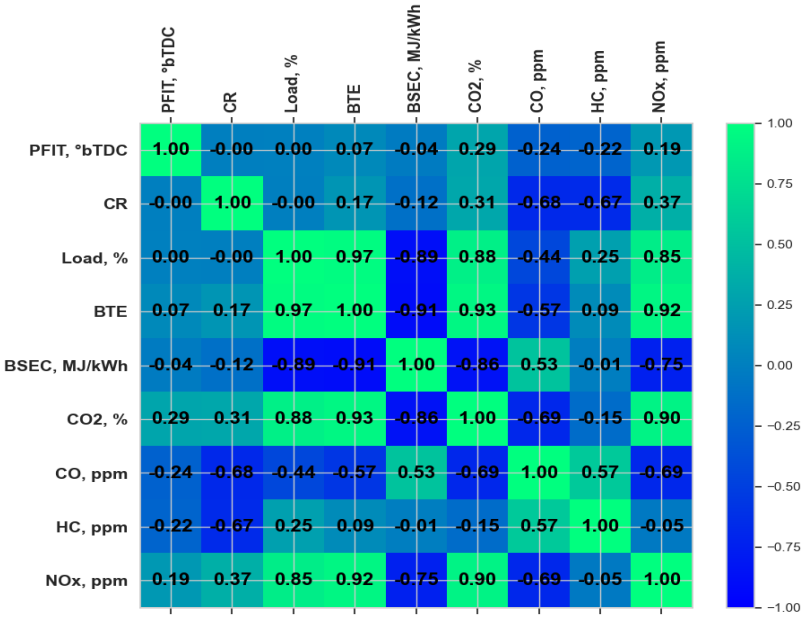

Fig. 3. Correlation heatmap

Table 2
Correlation matrix

	PFIT, °bTDC	CR	Load, %	BTE	BSEC, MJ/kWh	CO2, %	CO, ppm	HC, ppm	NOx, ppm
PFIT, °bTDC	1	0	0	0.07	-0.04	0.29	-0.24	-0.22	0.19
CR	0	1	0	0.17	-0.12	0.31	-0.68	-0.67	0.37
Load, %	0	0	1	0.97	-0.89	0.88	-0.44	0.25	0.85
BTE	0.07	0.17	0.97	1	-0.91	0.93	-0.57	0.09	0.92
BSEC, MJ/kWh	-0.04	-0.12	-0.89	-0.91	1	-0.86	0.53	-0.01	-0.75
CO2, %	0.29	0.31	0.88	0.93	-0.86	1	-0.69	-0.15	0.9
CO, ppm	-0.24	-0.68	-0.44	-0.57	0.53	-0.69	1	0.57	-0.69
HC, ppm	-0.22	-0.67	0.25	0.09	-0.01	-0.15	0.57	1	-0.05
NOx, ppm	0.19	0.37	0.85	0.92	-0.75	0.9	-0.69	-0.05	1

and enhanced efficiency. Conversely, BTE was inversely linked with Hydrocarbon (HC) emissions (-0.57) and Carbon Monoxide (CO) (-0.57). These inverse connections show that the generation of incomplete combustion byproducts like CO and HC declines as engine efficiency increases.

The Compression Ratio (CR) showed positive correlations with NOx (0.37) and CO_2 (0.31) emissions, implying that higher compression ratios contribute to increased NOx and CO_2 levels. Strong negative relationships between CR and CO (-0.68) and HC (-0.67) however indicate that increasing the compression ratio lowers these pollutants by improving combustion efficiency. With both BTE (-0.91) and NOx (-0.75), BSEC had considerable negative relationships. This suggests that lesser efficiency and more NOx emissions follow from increased energy use. On the other hand, a positive association between BSEC and CO emissions (0.53) shows that higher energy consump tion can raise CO emissions. These relationships highlight the complex trade-off required in optimizing biogas dual-fuel engines. Although better efficiency usually results in more NOx emissions, they may also lower incomplete combustion byproducts such as CO and HC.

4.2 Model development and evaluation

Using Python-based libraries including scikit-learn, pandas, numpy, and matplotlib, the model was developed. These models will be used to predict the performance of a biogas-diesel engine. The data will be gathered from lab-based testing in different operating settings. Data is divided into training and test sets using pandas and numpy. For predictive modeling, the researchers will then apply Lasso regression, Random Forest (RF), and support vector machines (SVM). The GridSearchCV approach was employed to tune hyperparameters and fit the models to the training data with scikit-learn. Employing metrics including Mean Squared Error (MSE) and R^2 score, model performance will be assessed, so allowing the researchers to compare their accuracy. At last, matplotlib will be used to show the outcomes, stressing the most successful method for biogas-diesel engine performance prediction and displaying the predictive capacity of every model.

4.2.1. Brake thermal efficiency models

Brake thermal efficiency *(*BTE) gauges how well fuel energy is turned into practical work. Higher BTE indicates effective combustion in biogas-diesel dual-fuel engines; its high methane concentration helps to improve combustion stability (Ambarita, 2017; Mohite *et al*., 2024). The prognostic models for BTE were developed using three ML approaches. Once the models were ready these were deployed for prediction, both during training and testing. The prediction results were evaluated using statistical methods as listed in Table 3. The results are also presented in a comparative manner as depicted in Figure 4a for the Lasso regression-based model, Figure 4b illustrates for SVM-based BTE model, while Figure 4c depicts RF based BTE model. The performance of various models varies greatly, as the data show in Table 3. With a test R^2 of 0.9338 and a test MSE value of 2.2173, lasso regression showed good performance. Although its test phase MAPE of 8.9651% shows a little degree of prediction error. With a lower test MSE of 1.1657 and a higher test R^2 of 0.9652, the SVM model beat the Lasso Regression. SVM's test MAPE of 9.8516% indicates, however, that while it increases prediction accuracy, it does not appreciably lower error margins. On the other hand, an RF-based model with a test MSE of 0.2074 and a good test R^2 of 0.9938, indicates almost flawless prediction accuracy. Thus, the RF came up as the best-performing model. Furthermore, RF had a low test MAPE of 3.0741%, indicating almost negligible prediction error. RF's ensemble learning method, which more successfully captures complicated patterns in the data than the linear correlations depicted by Lasso and the margin-based optimization of SVM, explains its better performance. Random Forest proved therefore the most reliable model for this purpose as it showed the highest performance in forecasting BTE.

Table 3
Statistical evaluation of the prediction results

Engine parameters	Model	Train MSE	Test MSE	Train R^2	Test R^2	Train MAPE, %	Test MAPE, %
BTE	Lasso Regression	1.4309	2.2173	0.9509	0.9338	7.8049	8.9651
	SVM	0	1.1657	1	0.9652	0	9.8516
	RF	0	0.2074	1	0.9938	0.0249	3.0741
BSEC	Lasso Regression	7.9442	7.7292	0.7611	0.8295	25.194	21.9774
	SVM	0	3.7904	1	0.9164	0	9.7179
	RF	0	1.2929	1	0.9715	0.0348	4.2242
CO_2	Lasso Regression	0.0961	0.0688	0.7377	0.8231	8.6037	8.2258
	SVM	0	0.0344	1	0.9115	0	6.2527
	RF	0	0	1	0.9821	0.0427	2.5801
CO	Lasso Regression	78.1325	85.5376	0.7076	0.6604	5.1313	5.2376
	SVM	0	97.25	1	0.6139	0	5.2794
	RF	0	41.8541	1	0.8339	0.0020	3.6099
HC	Lasso Regression	543.06	380.99	0.53	0.6097	10.57	10.4
	SVM	0	257.33	1	73.64	0	7.04
	RF	0	60.91	1	0.9376	0.01	3.28
NOx	Lasso Regression	781.4998	342.0372	0.9425	0.9760	14.3733	8.8571
	SVM	0	349.375	1	0.9754	0	6.5480
	RF	0	346.518	1	0.9756	0.0006	7.2056

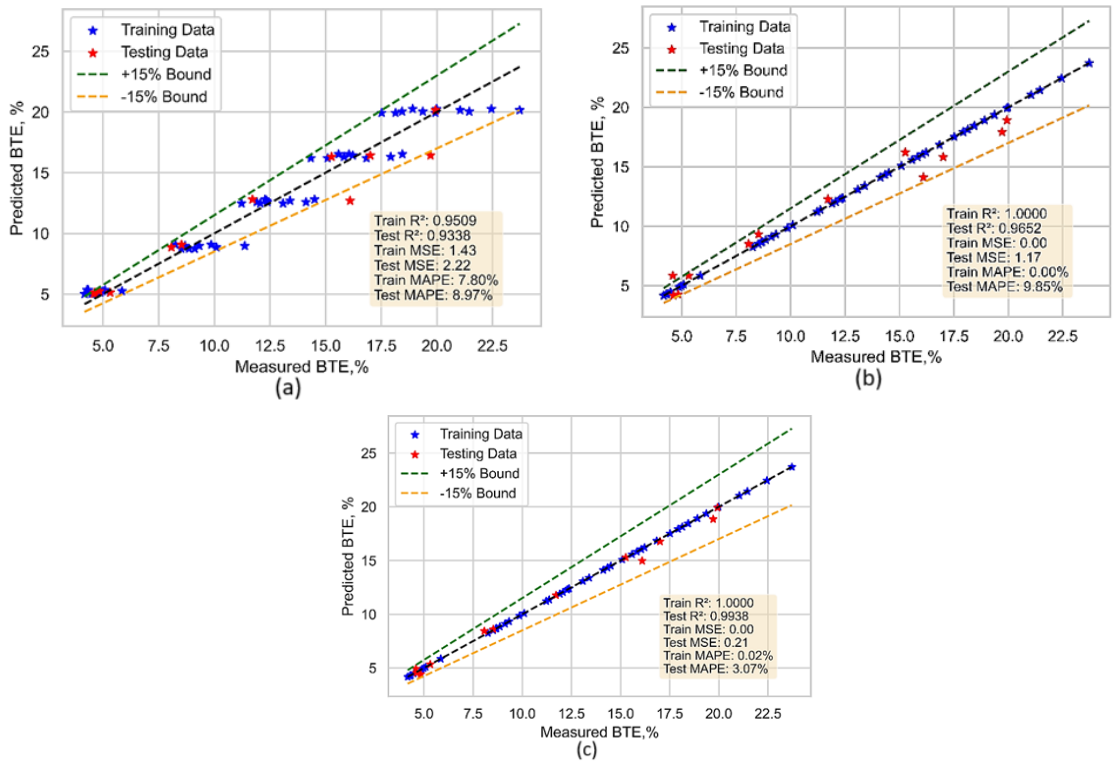

Fig. 4. BTE model's actual vs forecasted values for (a) Lasso regression (b) Support vector machine (c) Random Forest

4.2.2. Brake-specific energy consumption models

Brake-specific energy consumption (BSEC) stands for the energy used per unit of output from a job. Because of biogas's lower energy density, reduced BSEC in biogas-diesel dual-fuel engines results; yet exact injection techniques may help to maximize this. In the case of BSEC, three ML techniques were employed for the development of BSEC prognostic models. Once the models were ready, they were used for prediction—both for testing and training. Statistical approaches described in Table 3 were used to assess the prediction findings. Comparative presentation of the data is also shown in Figure 5a

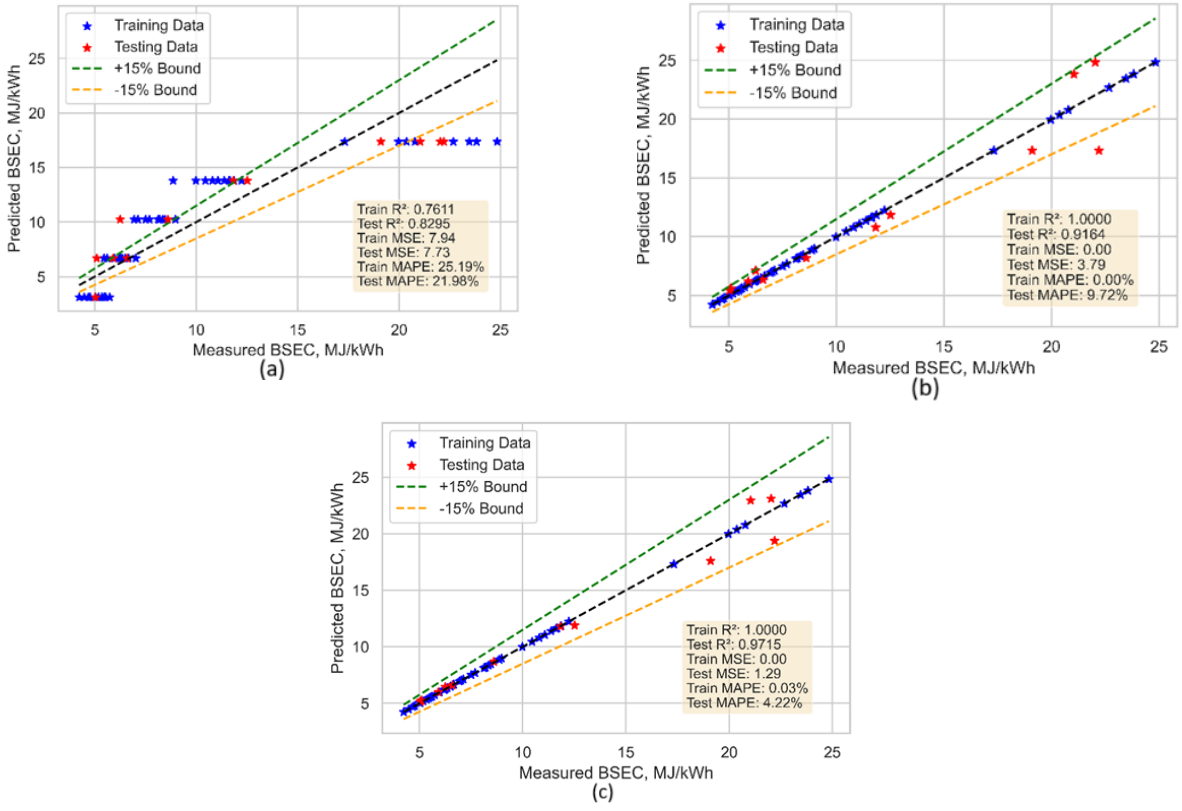

Fig. 5. BSEC model's actual vs forecasted values for (a) Lasso regression (b) Support vector machine (c) Random Forest

for the Lasso regression-based model, Figure 5b for SVM based BSEC model, and Figure 5c for RF based BSEC model. Given the facts shown in Table 3, different models perform somewhat differently. With a robust test MSE of 7.7292 and a test R^2 of 0.8295, lasso regression showed the poor performance showing low predictive accuracy. Its higher MAPE (21.9774%) during test phase also points to significant forecast variation. With a lower test MSE of 3.7904 and a higher test R^2 of 0.9164, SVM based BSEC model exhibited increased prediction accuracy in comparison with lasso regression. Showing better prediction consistency than Lasso, SVM also had lower test MAPE of 9.7179%. RF, on the other hand, was superior to the other two ML approaches with the lowest test MSE of 1.2929 and the best test R^2 of 0.9715. This indicates its great BSEC forecasting accuracy. With the lowest test MAPE, 4.2242%, RF also indicated the least prediction error. RF's capacity to manage complicated relationships between features and its resistance against overfitting help to explain its outstanding performance. Thus, in this work, Random Forest turned out to be the most successful model for BSEC prediction.

4.2.3. Carbon di-oxide emission models

Burning fuels derived from carbon bases produce CO_2 emissions. Because biogas has less carbon content than pure diesel, biogas-diesel dual-fuel engines produce less CO_2, hence lowering greenhouse gas emissions (Bora *et al.*, 2022; Kriaučiūnas *et al.*, 2021). The prognostic models developed using CO_2 emission data were used for prediction in the case of both testing and training. The prediction results were evaluated using statistical methods as listed in Table 3. The actual and model-predicted CO_2 emission data are compared as depicted in Figure 6a for the Lasso regression-based model, Figure 6b for SVM based CO_2 model, and Figure 6c for RF based CO_2 model. The data in Table 3 indicate the performance of different models. With a higher test MSE of 7.7292 and a test R^2 of 0.8295, lasso regression showed the lowest performance showing reduced predictive accuracy. Its higher test MAPE 21.9774% also points to significant forecast variation. With a lower test MSE of 3.7904 and a higher test R^2 of 0.9164, SVM exhibited increased prediction accuracy. Showing better prediction consistency than Lasso, SVM also lowered the test MAPE to 9.7179%.

With the lowest test MSE of 1.2929 and a high-test R^2 of 0.9715, RF based model exceeded both models. This indicates its great CO_2 emission forecasting accuracy. With the lowest test MAPE, 4.2242%, RF also indicated the least prediction error. RF's capacity to manage complicated relationships between features and its resistance against overfitting help to explain its outstanding performance. Thus, in this work, Random Forest turned out to be the best prediction model for CO_2 forecasting.

4.2.4. CO emission models

Incomplete combustion is one of the main causes of CO emissions in dual-fuel engines. In biogas-diesel dual-fuel engines, inadequate air-fuel mixing or poor injection time may raise CO emissions resulting from unburnt carbon monoxide from biogas combustion (Dobslaw *et al.*, 2019; Said *et al.*, 2022). In the case of CO emission data in the present study statistical outcomes listed in Table 3 were used to assess the prediction findings. For the Lasso regression-based model, Figure 7a shows the actual and model projected CO emission data; for SVM based CO model, Figure 7b; for RF based CO model, Figure 7c. It can be observed that most of the comparative data points are close to the best-fit line for all three models. Having a test MSE of 85.5376 and a test R^2 of 0.6604, lasso regression showed modest accuracy while nonetheless capturing data patterns but lacking precision. Reflecting a fair error rate, the model's test MAPE was 5.2376%. SVM underperformed in the test phase with a test MSE of 97.2500 and a lower test R^2 of

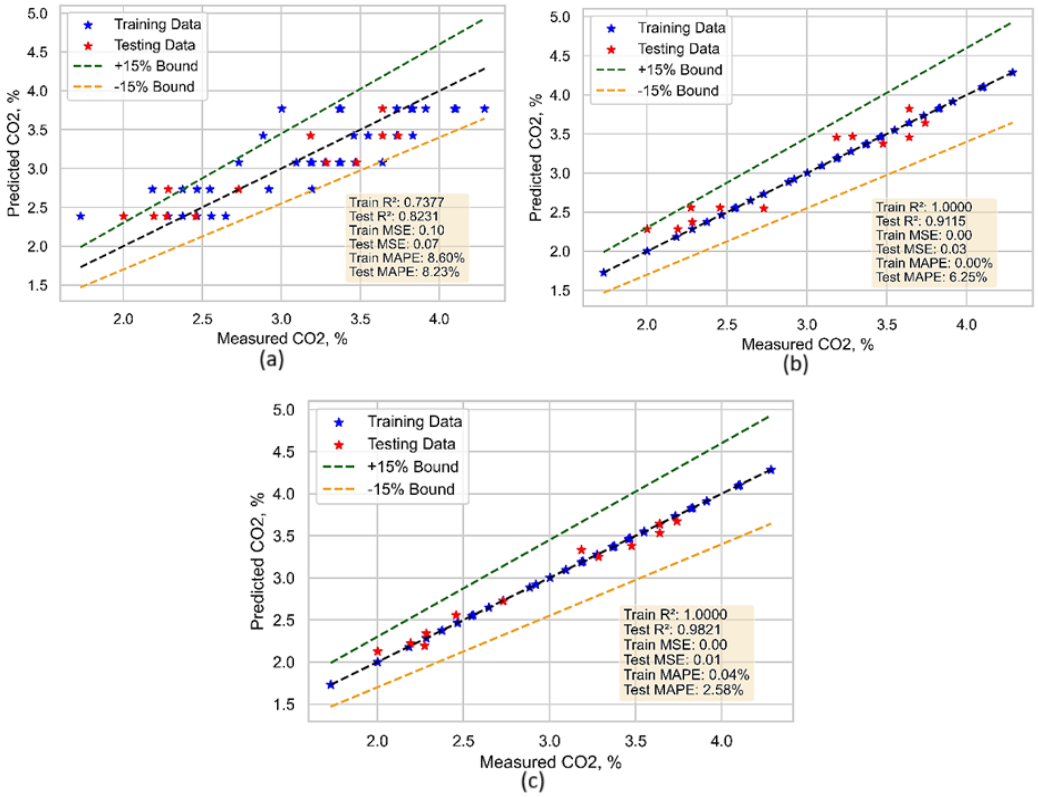

Fig. 6. CO_2 emission model's actual vs forecasted values for (a) Lasso regression (b) Support vector machine (c) Random Forest

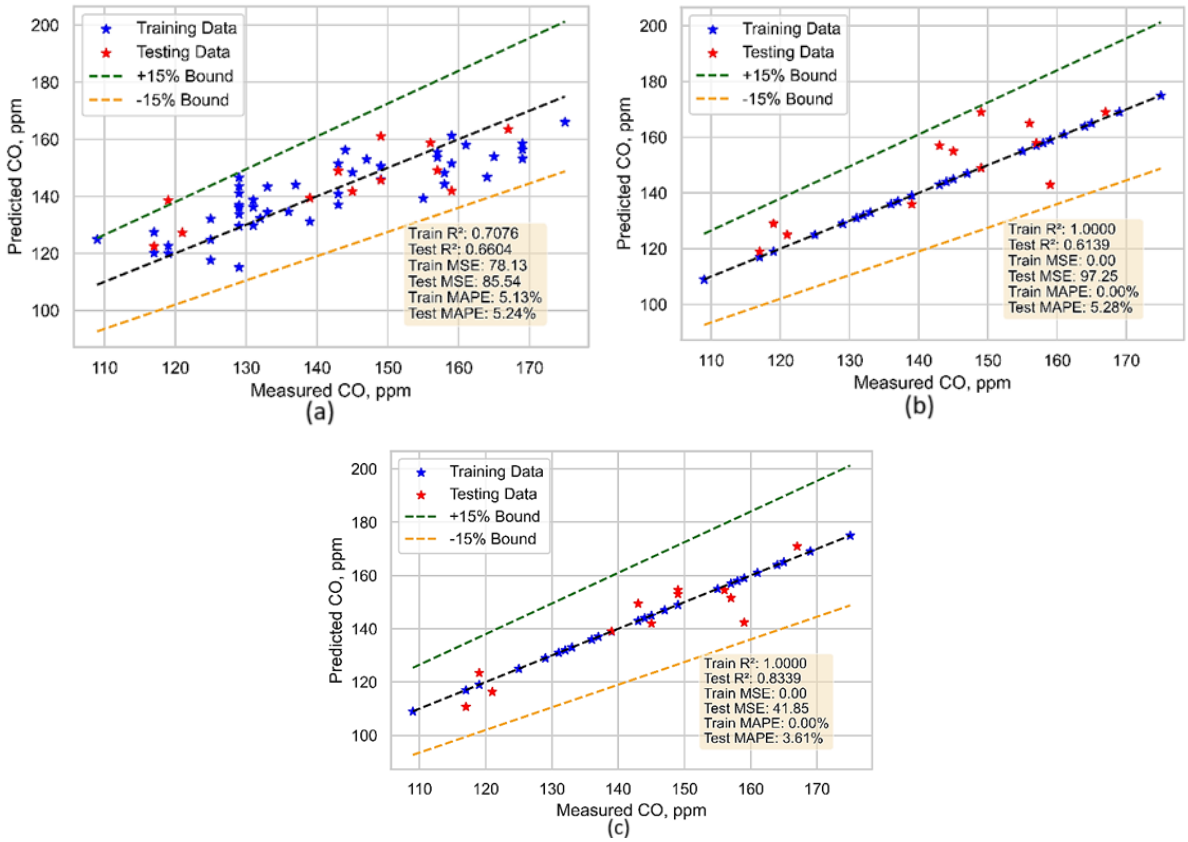

Fig. 7. CO emission model's actual vs forecasted values for (a) Lasso regression (b) Support vector machine (c) Random Forest

0.6139 while nonetheless obtaining a perfect train R^2 of 1. This points to overfitting, in which the model fails to generalize while fitting the training data precisely. With a far lower test, MSE of 41.8541 and a higher test R^2 of 0.8 339 Random Forest exceeded both models. Given its test MAPE of 3.6099%, RF was the most dependable model for estimating CO emissions in this context as it shows great accuracy.

4.2.5. HC emission models

Unburned fuels cause hydrocarbon (HC) emissions. Often resulting from poor injection timing or lower combustion temperatures, HC emissions in biogas-diesel dual-fuel engines may grow if biogas does not entirely burn. The models were developed using Lasso Regression, SVM, and RF for hydrocarbon (HC) emission prediction in a biogas diesel-powered dual-fuel engine (H. Sharma *et al.*, 2022). In the case of CO emission data in the current research, the prediction results are evaluated using Table 3 with statistical results. Figure 8a displays the actual and model-predicted CO emission data for the Lasso regression-based model; Figure 8b shows SVM based CO model; Figure 8c depicts RF based CO model. With a test R^2 of 0.6097 and a test MSE of 380.99, lasso regression has low predictive performance. The test MAPE of 10.4% indicates that it lacks accuracy even if it catches some of the data patterns exhibiting significant prediction errors. The training and testing performance of the SVM model shows a clear difference. Although the test R^2 of 73.64 points to overfitting, the ideal train R^2 of 1 is achieved. The model fails to generalize well even if it performs nicely during training. With overfitting, the test MAPE of 7.04% shows that the prediction errors remain somewhat modest. RF beats both versions noticeably. Its low test MSE of 60.91 and higher test R^2 of 0.9376 show great prediction accuracy. With a minimum prediction error of 3.28% the test MAPE reveals RF is the most dependable model for estimating HC emissions. Its better performance than Lasso Regression and SVM may be attributed in part to its capacity to manage complicated relationships in the data and prevent overfitting.

4.2.6. NOx emission models

High combustion temperatures create nitrogen oxides (NOx) (Nam *et al.*, 2024). The NOx for biogas-diesel dual-fuel mode remains on the lower side in comparison to diesel-powered engines, this is attributed to the lower combustion temperature of low-energy-density biogas. In the case of NOx emission data in this study, the statistical results are listed in Table 3. Figure 9a displays the actual and model-predicted CO emission data for the Lasso regression-based model; Figure 9b shows SVM based CO model; Figure 9c provides RF based CO model. With a test MSE of 342.0342 and a high R^2 of 0.9760 during the test phase, lasso regression showed good predictive accuracy. Although the model can record data patterns, its test MAPE of 8.8571% points to some potential for development in prediction accuracy. Closely mirroring Lasso Regression in terms of general accuracy, SVM showed comparable performance with a test MSE of 349.375 and a test R^2 of 0.9754. While preserving significant generalizing capacity, the test MAPE of 6.5450% shows somewhat greater accuracy in forecasting NOx emissions than Lasso Regression. Among the models, Random Forest (RF) came out as the best one. RF achieved outstanding accuracy with a test R^2 of 0.9756 and a test MSE of 346.518. RF is the most dependable model for NOx emissions prediction even although its test MAPE of 7.2056% was somewhat higher than that of SVM as its robustness in capturing intricate interactions and avoidance of overfitting. RF is the best option for this use overall because of its harmony of accuracy and precision.

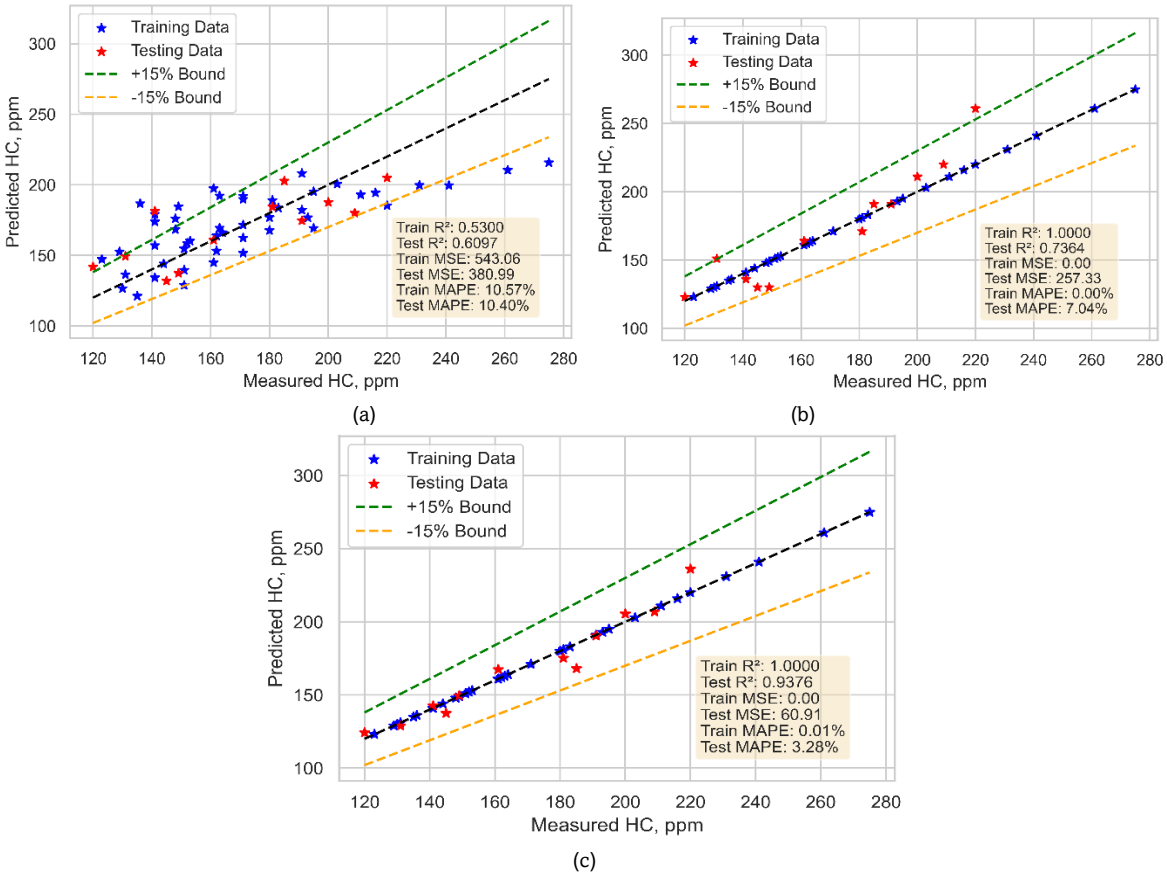

Fig. 8. HC emission model's actual vs forecasted values for (a) Lasso regression (b) Support vector machine (c) Random Forest

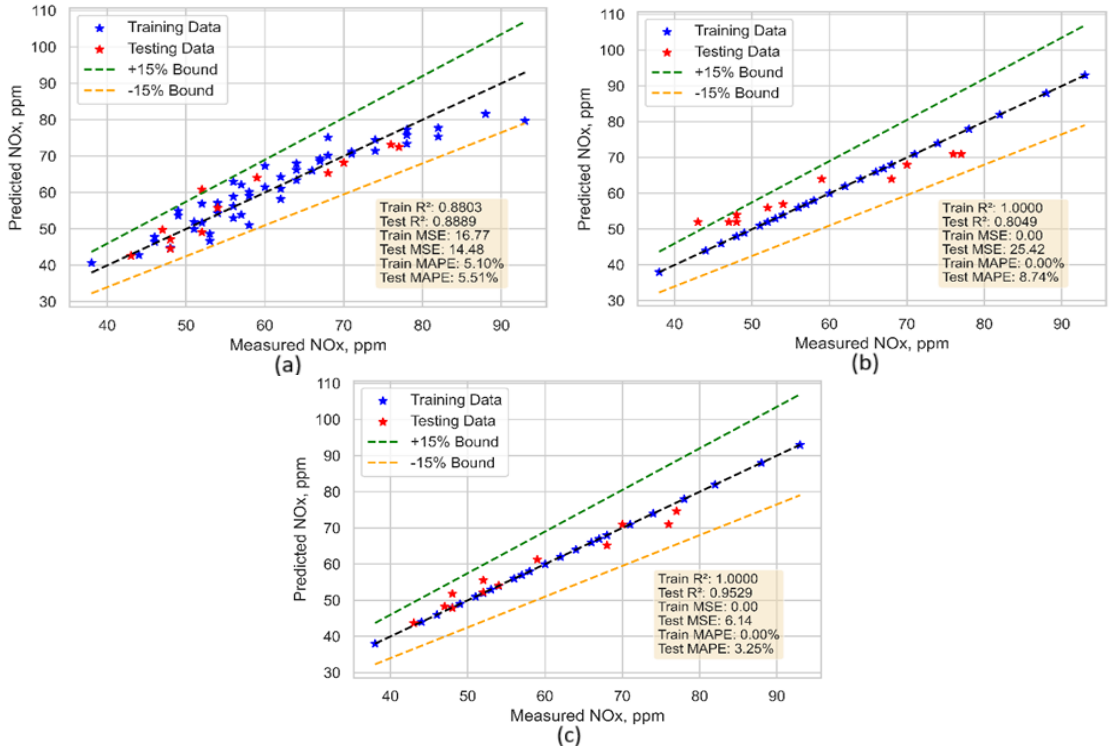

Fig. 9. NOx emission model's actual vs forecasted values for (a) Lasso regression (b) Support vector machine (c) Random Forest

4.2.7. Comparison of models

In the present study, the Taylor diagram was employed to compare the ML-based models. Taylor diagram graphically compared the output of the three machine learning models based on their predicted accuracy versus observed data, Lasso Regression, SVM, and RF. The graph depicts the correlation coefficient as radial contours and standard deviation along the

x and y axes, therefore enabling a complete evaluation of any model. The black star in this graphic stands in for the observed data and provides a reference point. The precision of any model is indicated by the closeness of its marker to this reference. In the case of the model training phase for BTE, the performance of the RF and SVM model almost replicates the measure values, hence indicating a strong predictive performance, as depicted in Figure 10a. However, Lasso regression was not that efficient in the case of the based model. On the other hand, the Taylor diagram for the test phase for the BTE model is depicted in Figure 10b. Represented as a blue circle, the RF model shows the lowest standard deviation and the closest correlation with the actual data. This suggests better forecasting accuracy. Showed as a red triangle, the SVM model likewise performs well but shows somewhat lower correlation and standard deviation. At last, the Lasso Regression shown by a blue square has the farthest location from the reference, therefore suggesting a worse performance than the other two models. This graph shows very clearly that in this specific situation, RF offers the most consistent forecasts. Based on this discussion, we can conclude the outcome of other models. The Taylor diagrams for training and testing phases in the case of BSEC models are depicted in Figure 11a and Figure 11b, respectively. In this case, it can be observed that in the training phase, RF and SVM performed excellently but Lasso Regression was not as good as it was away from observation (actual BSEC). However, in the case of the test phase, both SVM and Lasso regression performed poorly compared to RF.

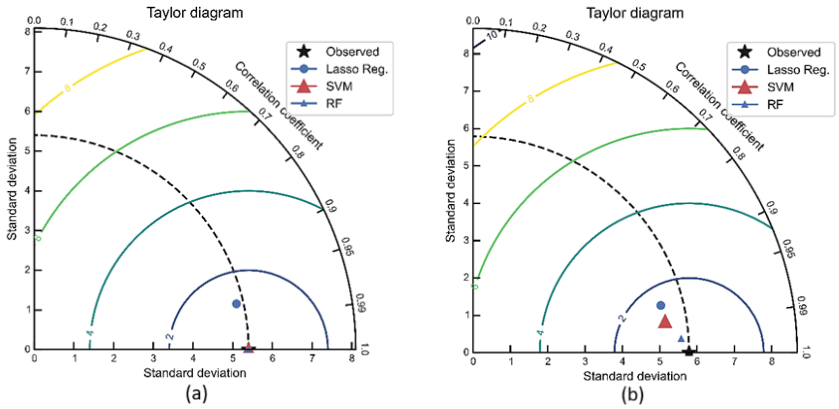

Fig. 10. Taylor's diagram for BTE model comparison for (a) training (b) testing phase

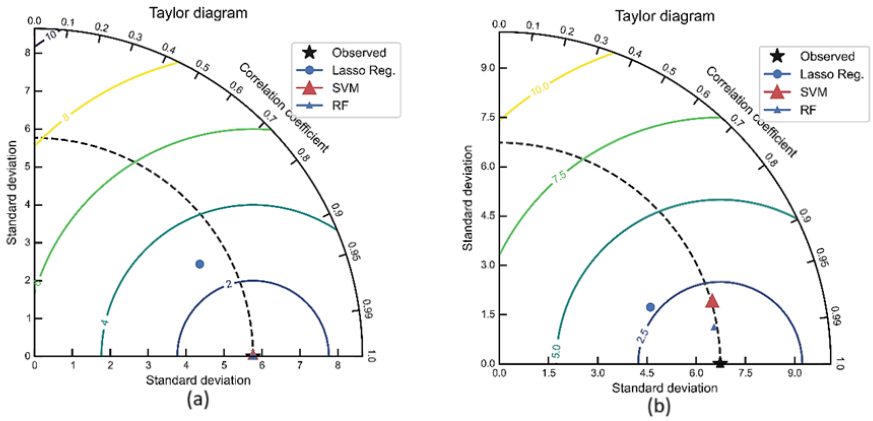

Fig. 11. Taylor's diagram for BSEC model comparison for (a) training (b) testing phase

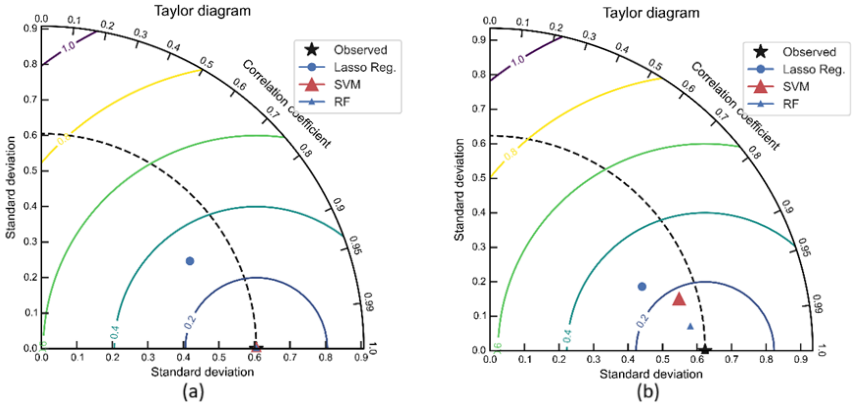

Fig. 12. Taylor's diagram for CO2 emission model comparison for (a) training (b) testing phase

Similarly, the Taylor diagrams were plotted for emission indices. Taylor diagrams for the CO2 emission model are depicted in Figure 12 while for CO emission the Taylor diagram is illustrated in Figure 13. The Taylor diagram shown for the CO_2 emission model as Figure 12a, for the model training phase shows that both RF as well as SVM predicted in a robust manner, while Lasso regression was comparatively poor. The Taylor diagram for the Test phase is plotted as Figure 12b illustrates that the performance of SVM was not that good while Lasso regression was again poorly performing. The RF-based CO2 emission model was the best-performing model in the model test phase. Figure 13a shows the Taylor diagram for the CO emission model over the model training period. Hereto, although Lasso regression performed badly, the RF and SVM both predicted strongly. Figure 13b displays the Taylor diagram during the Test phase, showing weak SVM performance and similarly poor Lasso regression performance. Throughout model testing, the RF-based CO emission model excelled.

The Taylor diagram for HC emission and NOx emission are depicted in Figure 14 and Figure 15 respectively. Under training and testing, the two Taylor diagrams in Figure 14a and Figure 14b show the performance of Lasso Regression, Support Vector Machine (SVM), and Random Forest (RF) models for HC emission. With the lowest standard deviation and maximum correlation to the observed data, Random Forest (blue circle) demonstrates the best performance in the training diagram (left), therefore showing accurate learning and powerful model training. With a greater standard deviation and somewhat lower

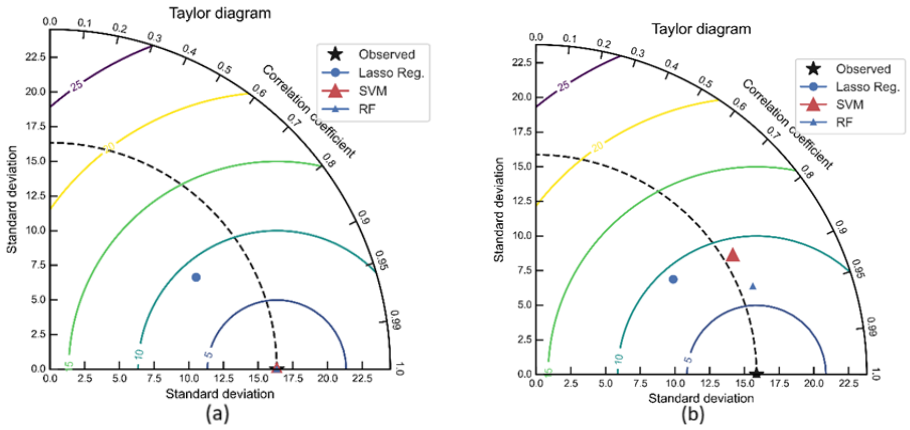

Fig. 13. Taylor's diagram for CO emission model comparison for (a) training (b) testing phase

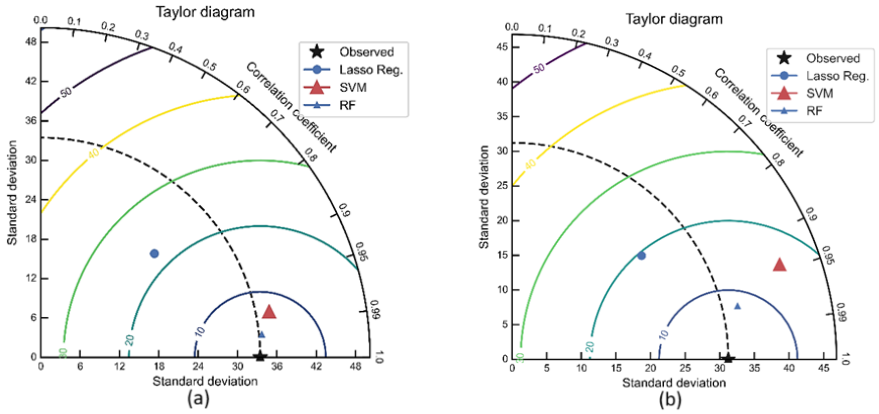

Fig. 14. Taylor's diagram for HC emission model comparison for (a) training (b) testing phase

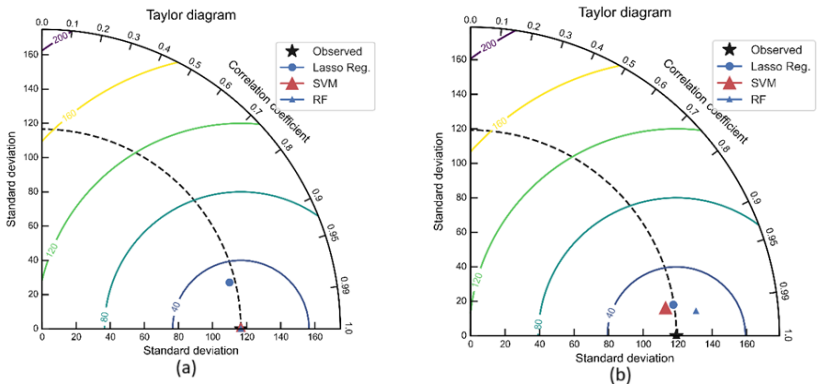

Fig. 15. Taylor's diagram for NOx emission model comparison for (a) training (b) testing phase

correlation, SVM (red triangle) performs well and suggests that its predictions are less exact than RF. With the lowest correlation and biggest standard deviation, lasso regression (blue square) indicates less efficient data acquisition and performs the poorest. Showing great generalization with low standard deviation and good correlation to the observed data, Random Forest keeps surpassing the other models in the testing diagram (right). While Lasso Regression once gives the lowest results and suffers from generalization and deviating much from the observed data, SVM follows with modest performance. Random Forest shows overall to be the most accurate and strong model in both stages; SVM and Lasso Regression follow.

For NOx emissions, Figure 15a and Figure 15b evaluate Lasso Regression, SVM, and RF models. With the lowest standard deviation and maximum correlation to the observed data, RF shown by the blue circle achieves the best performance in both training and testing phases, therefore suggesting accurate learning and excellent generalization. Given a larger standard deviation and somewhat reduced correlation, SVM shown as a red triangle performs well but is less exact than RF. With its blue square lowest correlation and biggest standard deviation, lasso regression which suggests ineffective data collection and poor generalization showcases the lowest performance. Though Lasso Regression trails behind, RF shows overall as the most strong and accurate model throughout both stages.

5. Conclusion

In conclusion, it is reported that biogas is a suitable alternative fuel to substitute diesel in a diesel engine. More so ever especially in dual-fuel mode with diesel as a pilot fuel, biogas offers good performance with lower emissions. This work simulated injection strategies for a waste-derived biogas-powered dual-fuel engine utilizing supervised machine learning methods including Random Forest, Lasso Regression, and Support Vector Machines (SVM). Mean Squared Error (MSE), R-squared (R^2), and Mean Absolute Percentage Error (MAPE) guided model assessments. Random Forest often showed better performance among the models across many criteria. For Brake Thermal Efficiency (BTE) forecasts, it obtained a test R^2 of 0.9938 and a low test MAPE of 3.0741%). With a test R^2 of 0.9715 and a test MAPE of 4.2242%, Random Forest once again exceeded other models in estimating Brake Specific Energy Consumption (BSEC). Random Forest demonstrated the best accuracy for CO_2 emissions with a test R^2 of 0.9821 and a test MAPE of 2.5801%. With a test R^2 of 0.8 339 and a test MAPE of 3.6099%, the model likewise performed remarkably in forecasting CO emissions. In the same vein, Random Forest outperformed Linear Regression in NOx emission with a test R^2 of 0.9756 and a test MAPE of 7.2056%. Random Forest was demonstrated to be the most consistent model. This work emphasizes how well machine learning especially Random Forest can forecast biogas dual-fuel engine performance and emissions.

References

Abdullah, B., Syed Muhammad, S.A.F., Shokravi, Z., Ismail, S., Kassim, K.A., Mahmood, A.N., Aziz, M.M.A., 2019. Fourth generation biofuel: A review on risks and mitigation strategies. *Renew. Sustain. Energy Rev.* 107, 37–50. https://doi.org/10.1016/j.rser.2019.02.018

Ahmad, A., Yadav, A.K., Hasan, S., 2024a. Enhanced production of methane enriched biogas through intensified co-digestion process and its effective utilization in a biodiesel/biohydrogen fueled engine with duel injection strategies: ML-RSM based an efficient optimization approach. *Int. J. Hydrogen Energy* 65, 671–686. https://doi.org/10.1016/J.IJHYDENE.2024.04.059

Ahmad, A., Yadav, A.K., Singh, A., Singh, D.K., 2024b. A comprehensive machine learning-coupled response surface methodology approach for predictive modeling and optimization of biogas potential in anaerobic Co-digestion of organic waste. *Biomass and Bioenergy* 180, 106995. https://doi.org/10.1016/j.biombioe.2023.106995

Ambarita, H., 2017. Performance and emission characteristics of a small diesel engine run in dual-fuel (diesel-biogas) mode. *Case Stud. Therm. Eng.* 10, 179–191. https://doi.org/10.1016/j.csite.2017.06.003

Ardolino, F., Cardamone, G.F., Parrillo, F., Arena, U., 2021. Biogas-to-biomethane upgrading: A comparative review and assessment in a life cycle perspective. *Renew. Sustain. Energy Rev.* 139, 110588. https://doi.org/10.1016/j.rser.2020.110588

Awe, O.W., Zhao, Y., Nzihou, A., Minh, D.P., Lyczko, N., 2017. A Review of Biogas Utilisation, Purification and Upgrading Technologies. *Waste and Biomass Valorization* 8, 267–283. https://doi.org/10.1007/s12649-016-9826-4

Ayyıldız, E., Murat, M., 2024. A lasso regression-based forecasting model for daily gasoline consumption: Türkiye Case. *Turkish J. Eng.* 8, 162–174. https://doi.org/10.31127/TUJE.1354501

Bhatt, A.N., Shrivastava, N., 2021. Application of Artificial Neural Network for Internal Combustion Engines: A State of the Art Review. Arch. *Comput. Methods Eng.* 1–23.

Bietresato, M., Calcante, A., Mazzetto, F., 2015. A neural network approach for indirectly estimating farm tractors engine performances. *Fuel* 143, 144–154. https://doi.org/10.1016/J.FUEL.2014.11.019

Bora, B.J., Dai Tran, T., Prasad Shadangi, K., Sharma, P., Said, Z., Kalita, P., Buradi, A., Nhanh Nguyen, V., Niyas, H., Tuan Pham, M., Thanh Nguyen Le, C., Dung Tran, V., Phuong Nguyen, X., 2022. Improving combustion and emission characteristics of a biogas/biodiesel-powered dual-fuel diesel engine through trade-off analysis of operation parameters using response surface methodology. *Sustain. Energy Technol. Assessments 53*, 102455. https://doi.org/10.1016/j.seta.2022.102455

Bora, B.J., Saha, U.K., 2016. Optimisation of injection timing and compression ratio of a raw biogas powered dual fuel diesel engine. *Appl. Therm. Eng.* 92, 111–121. https://doi.org/10.1016/j.applthermaleng.2015.08.111

Boretti, A., 2019. Advantages and Disadvantages of Diesel Single and Dual-Fuel Engines. *Front. Mech. Eng.* 5. https://doi.org/10.3389/fmech.2019.00064

Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32. https://doi.org/10.1023/A:1010933404324

Bui, V.G., Bui, T.M.T., Hoang, A.T., Nižetić, S., Nguyen Thi, T.X., Vo, A.V., 2021. Hydrogen-Enriched Biogas Premixed Charge Combustion and Emissions in Direct Injection and Indirect Injection Diesel Dual Fueled Engines: A Comparative Study. *J. Energy Resour. Technol.* 143. https://doi.org/10.1115/1.4051574

Bui, V.G., Bui, T.M.T., Nguyen, M.T., Bui, V.H., Do, P.N., Tran, N.A.H., Le, T.T., Hoang, A.T., 2024. Enhancing the performance of syngas-diesel dual-fuel engines by optimizing injection regimes: From comparative analysis to control strategy proposal. *Process Saf. Environ. Prot.* 186, 1034–1052. https://doi.org/10.1016/j.psep.2024.04.042

Bui, V.G., Bui, T.M.T., Tran, V.N., Huang, Z., Hoang, A.T., Tarelko, W., Bui, V.H., Pham, X.M., Nguyen, P.Q.P., 2023. Flexible syngas-biogas-hydrogen fueling spark-ignition engine behaviors with optimized fuel compositions and control parameters. *Int. J. Hydrogen Energy* 48, 6722–6737. https://doi.org/10.1016/j.ijhydene.2022.09.133

Bui, V.G., Tu Bui, T.M., Ong, H.C., Nižetić, S., Bui, V.H., Xuan Nguyen, T.T., Atabani, A.E., Štěpanec, L., Phu Pham, L.H., Hoang, A.T., 2022. Optimizing operation parameters of a spark-ignition engine fueled with biogas-hydrogen blend integrated into biomass-solar hybrid renewable energy system. *Energy* 252, 124052. https://doi.org/10.1016/j.energy.2022.124052

Changxiong, L., Hu, Y., Yang, Z., Guo, H., 2023. Experimental Study of Fuel Combustion and Emission Characteristics of Marine Diesel Engines Using Advanced Fuels. *Polish Marit. Res.* 30, 48–58. https://doi.org/10.2478/pomr-2023-0038

Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20, 273–297. https://doi.org/10.1007/BF00994018

Cunanan, C., Tran, M.-K., Lee, Y., Kwok, S., Leung, V., Fowler, M., 2021. A Review of Heavy-Duty Vehicle Powertrain Technologies: Diesel Engine Vehicles, Battery Electric Vehicles, and Hydrogen Fuel Cell Electric Vehicles. Clean Technol. 3, 474–489. https://doi.org/10.3390/cleantechnol3020028

Czekała, W., 2022. Biogas as a Sustainable and Renewable Energy Source. pp. 201–214. https://doi.org/10.1007/978-981-16-8747-1_10

Dobslaw, D., Engesser, K.-H., Störk, H., Gerl, T., 2019. Low-cost process for emission abatement of biogas internal combustion engines. J. Clean. Prod. 227, 1079–1092. https://doi.org/10.1016/j.jclepro.2019.04.258

E, J., Pham, M., Zhao, D., Deng, Y., Le, D., Zuo, W., Zhu, H., Liu, T., Peng, Q., Zhang, Z., 2017. Effect of different technologies on combustion and emissions of the diesel engine fueled with biodiesel: A review. *Renew. Sustain. Energy Rev.* 80, 620–647. https://doi.org/10.1016/j.rser.2017.05.250

Elvidge, S., Angling, M.J., Nava, B., 2014. On the use of modified Taylor diagrams to compare ionospheric assimilation models. Radio Sci. 49, 737–745. https://doi.org/10.1002/2014RS005435

Feroskhan, M., Ismail, S., 2017. A review on the purification and use of biogas in compression ignition engines. *Int. J. Automot. Mech. Eng.* 14, 4383–4400. https://doi.org/10.15282/ijame.14.3.2017.1.0348

Feroskhan, M., Ismail, S., Reddy, M.G., Sai Teja, A., 2018. Effects of charge preheating on the performance of a biogas-diesel dual fuel CI engine. *Eng. Sci. Technol. an Int. J.* 21, 330–337. https://doi.org/10.1016/j.jestch.2018.04.001

Fransiscus, Y., Simangunsong, T.L., 2021. Anaerobic Digestion of Industrial Tempeh Wastewater with Sludge from Cow Manure Biogas Digester as Inoculum: Effect of F/M Ratio on the Methane Production. *Int. J. Adv. Sci. Eng. Inf. Technol.* 11, 1007–1013. https://doi.org/10.18517/ijaseit.11.3.11846

García-Nieto, P.J., García-Gonzalo, E., José, •, Paredes-Sá Nchez, P., n.d. Prediction of the critical temperature of a superconductor by using the WOA/MARS, Ridge, Lasso and Elastic-net machine learning techniques. *Neural Comput. Appl.* 33. https://doi.org/10.1007/s00521-021-06304-z

Gholizadeh, M., Jamei, M., Ahmadianfar, I., Pourrajab, R., 2020. Prediction of nanofluids viscosity using random forest (RF) approach. Chemom. Intell. *Lab. Syst.* 201, 104010. https://doi.org/10.1016/J.CHEMOLAB.2020.104010

Goyal, D., Goyal, T., Mahla, S.K., Goga, G., Dhir, A., Balasubramanian, D., Hoang, A.T., Wae-Hayee, M., Josephin, J.S.F., Sonthalia, A., Varuvel, E.G., Brindhadevi, K., 2023. Application of Taguchi design in optimization of performance and emissions characteristics of n-butanol/diesel/biogas under dual fuel mode. *Fuel* 338, 127246. https://doi.org/10.1016/j.fuel.2022.127246

Hao, D., Mehra, R.K., Luo, S., Nie, Z., Ren, X., Fanhua, M., 2020. Experimental study of hydrogen enriched compressed natural gas (HCNG) engine and application of support vector machine (SVM) on prediction of engine performance at specific condition. *Int. J. Hydrogen Energy* 45, 5309–5325. https://doi.org/10.1016/j.ijhydene.2019.04.039

Haque, R., Quek, A., Ting, C.-Y., Goh, H.-N., Hasan, M.R., 2024. Classification Techniques Using Machine Learning for Graduate Student Employability Predictions. Int. J. Adv. Sci. Eng. Inf. Technol. 14, 45–56. https://doi.org/10.18517/ijaseit.14.1.19549

Hidayanti, F., Wati, E.K., Sumardi, A.H., 2021. Implementation of Automatic Monitoring and Control System in Balloon Digester Type's Biogas Plant from Tofu Waste Using Ultrasonic Distance Sensor. *Int. J. Adv. Sci. Eng. Inf. Technol.* 11, 1463–1467. https://doi.org/10.18517/ijaseit.11.4.11004

Hoang, A.T., 2021. Combustion behavior, performance and emission characteristics of diesel engine fuelled with biodiesel containing cerium oxide nanoparticles: A review. *Fuel Process. Technol.* 218, 106840. https://doi.org/10.1016/j.fuproc.2021.106840

Hoang, A.T., 2018. Waste heat recovery from diesel engines based on Organic Rankine Cycle. *Appl. Energy* 231, 138–166. https://doi.org/10.1016/j.apenergy.2018.09.022

Hoang, A.T., Goldfarb, J.L., Foley, A.M., Lichtfouse, E., Kumar, M., Xiao, L., Ahmed, S.F., Said, Z., Luque, R., Bui, V.G., Nguyen, X.P., 2022. Production of biochar from crop residues and its application for anaerobic digestion. *Bioresour. Technol.* 363, 127970. https://doi.org/10.1016/j.biortech.2022.127970

Hoang, A.T., Pandey, A., Martinez De Osés, F.J., Chen, W.-H., Said, Z., Ng, K.H., Ağbulut, Ü., Tarełko, W., Ölçer, A.I., Nguyen, X.P., 2023. Technological solutions for boosting hydrogen role in decarbonization strategies and net-zero goals of world shipping: Challenges and perspectives. *Renew. Sustain. Energy Rev.* 188, 113790. https://doi.org/10.1016/j.rser.2023.113790

Jamei, M., Pourrajab, R., Ahmadianfar, I., Noghrehabadi, A., 2020. Accurate prediction of thermal conductivity of ethylene glycol-based hybrid nanofluids using artificial intelligence techniques. *Int. Commun. Heat Mass Transf.* 116, 104624. https://doi.org/10.1016/J.ICHEATMASSTRANSFER.2020.104624

Jamei, M., Sharma, P., Ali, M., Bora, B.J., Malik, A., Paramasivam, P., Farooque, A.A., Abdulla, S., 2024. Application of an explainable glass-box machine learning approach for prognostic analysis of a biogas-powered small agriculture engine. *Energy* 288, 129862. https://doi.org/10.1016/j.energy.2023.129862

Kanti, P.K., Sharma, P., Maiya, M.P., Sharma, K.V., 2023. The stability and thermophysical properties of Al2O3-graphene oxide hybrid nanofluids for solar energy applications: Application of robust autoregressive modern machine learning technique. *Sol. Energy Mater. Sol. Cells* 253, 112207. https://doi.org/10.1016/j.solmat.2023.112207

Kapoor, R., Ghosh, P., Tyagi, B., Vijay, V.K., Vijay, V., Thakur, I.S., Kamyab, H., Nguyen, D.D., Kumar, A., 2020. Advances in biogas valorization and utilization systems: A comprehensive review. *J. Clean. Prod.* 273, 123052.

Karne, H., Mahajan, U., Ketkar, U., Kohade, A., Khadilkar, P., Mishra, A., 2023. A review on biogas upgradation systems. *Mater. Today Proc.* 72, 775–786. https://doi.org/10.1016/J.MATPR.2022.09.015

Khayum, N., Anbarasu, S., Murugan, S., 2020. Combined effect of fuel injecting timing and nozzle opening pressure of a biogas-biodiesel fuelled diesel engine. *Fuel* 262, 116505. https://doi.org/10.1016/j.fuel.2019.116505

Kriaučiūnas, D., Pukalskas, S., Rimkus, A., Barta, D., 2021. Analysis of the Influence of CO2 Concentration on a Spark Ignition Engine Fueled with Biogas. *Appl. Sci.* 11, 6379. https://doi.org/10.3390/app11146379

Kumar Kanti, P., Sharma, P., Sharma, K.V., Maiya, M.P., 2023. The effect of pH on stability and thermal performance of graphene oxide and copper oxide hybrid nanofluids for heat transfer applications: Application of novel machine learning technique. *J. Energy Chem.* 82, 359–374. https://doi.org/10.1016/j.jechem.2023.04.001

Li, Y., Yang, R., Wang, X., Zhu, J., Song, N., 2023. Carbon Price Combination Forecasting Model Based on Lasso Regression and Optimal Integration. Sustain. 2023, Vol. 15, Page 9354 15, 9354. https://doi.org/10.3390/SU15129354

Lohani, S.P., Dhungana, B., Horn, H., Khatiwada, D., 2021. Small-scale biogas technology and clean cooking fuel: Assessing the potential and links with SDGs in low-income countries – A case study of Nepal. Sustain. *Energy Technol. Assessments* 46, 101301. https://doi.org/10.1016/J.SETA.2021.101301

Mahla, S.K., Singla, V., Sandhu, S.S., Dhir, A., 2018. Studies on biogas-fuelled compression ignition engine under dual fuel mode. *Environ. Sci. Pollut. Res.* 25, 9722–9729. https://doi.org/10.1007/S11356-018-1247-4/METRICS

Mao, C., Feng, Y., Wang, X., Ren, G., 2015. Review on research achievements of biogas from anaerobic digestion. *Renew. Sustain. Energy Rev.* 45, 540–555. https://doi.org/10.1016/j.rser.2015.02.032

Meenal, R., Selvakumar, A.I., 2018. Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters. *Renew. Energy* 121, 324–343. https://doi.org/10.1016/J.RENENE.2017.12.005

Meng, X., Zhou, Y., Yang, T., Long, W., Bi, M., Tian, J., Lee, C.F.F., 2020. An experimental investigation of a dual-fuel engine by using bio-fuel as the additive. *Renew. Energy* 147, 2238–2249. https://doi.org/10.1016/J.RENENE.2019.10.023

Mohammad, A., Rezaei, R., Hayduk, C., Delebinski, T., Shahpouri, S., Shahbakhti, M., 2022. Physical-oriented and machine learning-based emission modeling in a diesel compression ignition engine: Dimensionality reduction and regression. 24, 904–918. https://doi.org/10.1177/14680874211070736

Mohite, A., Bora, B.J., Sharma, P., Sarıdemir, S., Mallick, D., Sunil, S., Ağbulut, Ü., 2024. Performance enhancement and emission control through adjustment of operating parameters of a biogas-biodiesel dual fuel diesel engine: An experimental and statistical study with biogas as a hydrogen carrier. *Int. J. Hydrogen Energy* 52, 752–764.

Naghavi, R., Abdoli, M.A., Karbasi, A., Adl, M., 2020. Improving the Quantity and Quality of Biogas Production in Tehran Anaerobic Digestion Power Plant by Application of Materials Recirculation Technique. *Int. J. Renew. Energy Dev.* 9, 167–175. https://doi.org/10.14710/ijred.9.2.167-175

Najafi, G., Ghobadian, B., Moosavian, A., Yusaf, T., Mamat, R., Kettner, M., Azmi, W.H., 2016. SVM and ANFIS for prediction of performance and exhaust emissions of a SI engine with gasoline–ethanol blended fuels. *Appl. Therm. Eng.* 95, 186–203. https://doi.org/10.1016/j.applthermaleng.2015.11.009

Nam, D., Jafrin, A., Johnson, T., 2024. A Simulation Study on a Premixed-charge Compression Ignition Mode-based Engine Using a Blend of Biodiesel / Diesel Fuel under a Split Injection Strategy. *Int. J. Adv. Sci. Eng. Inf. Technol.* 14, 451–471.

Nguyen-Thi, T.X., Bui, T.M.T., 2023. Effects of Injection Strategies on Mixture Formation and Combustion in a Spark-Ignition Engine Fueled with Syngas-Biogas-Hydrogen. *Int. J. Renew. Energy Dev.* 12, 118–128. https://doi.org/10.14710/ijred.2023.49368

Nguyen, V.N., Tarełko, W., Sharma, P., El-Shafay, A.S., Chen, W.-H., Nguyen, P.Q.P., Nguyen, X.P., Hoang, A.T., 2024. Potential of Explainable Artificial Intelligence in Advancing Renewable Energy: Challenges and Prospects. Energy & Fuels 38, 1692–1712. https://doi.org/10.1021/acs.energyfuels.3c04343

Nisa, Z.K., Pradipta, A.G., Sholikah, L.N., Pratama, B.F., Prihanantya, A.S., Ngadisih, Susanto, S., Arif, S.S., 2023. Recognition of Agricultural Land-Use Change with Machine Learning-Based for Regional Food Security Assessment in Kulon Progo Plains Area. Int. J. Adv. Sci. Eng. Inf. Technol. 13, 54–61. https://doi.org/10.18517/ijaseit.13.1.16550

Onukwuli, D.O., Esonye, C., Ofoefule, A.U., Eyisi, R., 2021. 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. 125, 153–167. https://doi.org/10.1016/J.JTICE.2021.06.012

Parsaee, M., Kiani Deh Kiani, M., Karimi, K., 2019. A review of biogas production from sugarcane vinasse. Biomass and Bioenergy 122, 117–125. https://doi.org/10.1016/J.BIOMBIOE.2019.01.034

Patnaik, S., Khatri, N., Rene, E.R., 2024. Artificial neural networks-based performance and emission characteristics prediction of compression ignition engines powered by blends of biodiesel derived from waste cooking oil. *Fuel* 370, 131806. https://doi.org/10.1016/J.FUEL.2024.131806

Pisner, D.A., Schnyer, D.M., 2020. Support vector machine, in: Machine Learning. Elsevier, pp. 101–121. https://doi.org/10.1016/B978-0-12-815739-8.00006-7

Prabhu, A.V., Avinash, A., Brindhadevi, K., Pugazhendhi, A., 2021. Performance and emission evaluation of dual fuel CI engine using preheated biogas-air mixture. *Sci. Total Environ.* 754, 142389. https://doi.org/10.1016/J.SCITOTENV.2020.142389

Puri, D., Nalbalwar, S., Nandgaonkar, A., Rajput, J., Wagh, A., 2023. Identification of Alzheimer's Disease Using Novel Dual Decomposition Technique and Machine Learning Algorithms from EEG Signals. *Int. J. Adv. Sci. Eng. Inf. Technol.* 13, 658–665. https://doi.org/10.18517/ijaseit.13.2.18252

Rahman, H., Nehemia, A., Astuti, H.P., 2023. Investigating the potential of avocado seeds for bioethanol production: A study on boiled water delignification pretreatment. *Int. J. Renew. Energy Dev.* 12(4). https://doi.org/10.14710/ijred.2023.52532 .

Ramalingam, K., Vellaiyan, S., Venkatesan, E.P., Khan, S.A., Mahmoud, Z., Saleel, C.A., 2023. Challenges and Opportunities of Low Viscous Biofuel—A Prospective Review. *ACS Omega* 8, 16545–16560. https://doi.org/10.1021/acsomega.3c00387

Rao, A., Liu, Y., Ma, F., 2022. Study of NOx emission for hydrogen enriched compressed natural along with exhaust gas recirculation in spark ignition engine by Zeldovich' mechanism, support vector machine and regression correlation. *Fuel* 318, 123577. https://doi.org/10.1016/j.fuel.2022.123577

Rocha-Meneses, L., Luna-delRisco, M., González, C.A., Moncada, S.V., Moreno, A., Sierra-Del Rio, J., Castillo-Meza, L.E., 2023. An Overview of the Socio-Economic, Technological, and Environmental Opportunities and Challenges for Renewable Energy Generation from Residual Biomass: A Case Study of Biogas Production in Colombia. Energies 2023, Vol. 16, Page 5901 16, 5901. https://doi.org/10.3390/EN16165901

Runyowa, N.T., Fourie, W., 2021. Responsible Leadership and the Implementation of SDG 7: The Case of the UNDP Botswana Biogas Project. *Sustain. Dev. Goals Ser. Part F2673*, 91–105. https://doi.org/10.1007/978-3-030-70952-5_7

Sahoo, B.B., Sahoo, N., Saha, U.K., 2009. Effect of engine parameters and type of gaseous fuel on the performance of dual-fuel gas diesel engines—A critical review. *Renew. Sustain. Energy Rev.* 13, 1151–1184. https://doi.org/10.1016/j.rser.2008.08.003

Said, Z., Sharma, P., Bora, B.J., Bui, T.A.E., Nguyen, D.T., Dinh, X.T., Nguyen, X.P., 2022. Modeling-optimization of performance and emission characteristics of dual-fuel engine powered with pilot diesel and agricultural-food waste-derived biogas. *Int. J. Hydrogen Energy*.

Sakthivel, R., Ramesh, K., Purnachandran, R., Mohamed Shameer, P., 2018. A review on the properties, performance and emission aspects of the third generation biodiesels. *Renew. Sustain. Energy Rev.* 82, 2970–2992. https://doi.org/10.1016/j.rser.2017.10.037

Schonlau, M., Zou, R.Y., 2020. The random forest algorithm for statistical learning. Stata J. Promot. Commun. Stat. Stata 20, 3–29. https://doi.org/10.1177/1536867X20909688

Serbin, S., Burunsuz, K., Chen, D., Kowalski, J., 2022. Investigation of the Characteristics of a Low-Emission Gas Turbine Combustion Chamber Operating on a Mixture of Natural Gas and Hydrogen. Polish Marit. Res. 29, 64–76. https://doi.org/10.2478/pomr-2022-0018

Sharma, H., Mahla, S.K., Dhir, A., 2022. Effect of utilization of hydrogen-rich reformed biogas on the performance and emission characteristics of common rail diesel engine. *Int. J. Hydrogen Energy* 47, 10409–10419. https://doi.org/10.1016/j.ijhydene.2022.01.073

Sharma, P.;, Bora, B.J.A., Sharma, Prabhakar, Bora, B.J., 2022. A Review of Modern Machine Learning Techniques in the Prediction of Remaining Useful Life of Lithium-Ion Batteries. *Batter.* 13(9), 13. https://doi.org/10.3390/BATTERIES9010013

Sharma, P., Balasubramanian, D., Thanh Khai, C., Papla Venugopal, I., Alruqi, M., Josephin JS, F., Sonthalia, A., Geo Varuvel, E., Khalife, E., Ravikumar, R., Wae-Hayee, M., 2023a. Enhancing the performance of renewable biogas powered engine employing oxyhydrogen: Optimization with desirability and D-optimal design. *Fuel* 341, 127575. https://doi.org/10.1016/j.fuel.2023.127575

Sharma, P., Bora, B.J., 2023. Modeling and optimization of a CI engine running on producer gas fortified with oxyhydrogen. Energy 270, 126909. https://doi.org/10.1016/j.energy.2023.126909

Sharma, P., Jain, A., Bora, B.J., Balakrishnan, D., Show, P.L., Ramaraj, R., Ağbulut, Ü., Khoo, K.S., 2023b. Application of modern approaches to the synthesis of biohydrogen from organic waste. *Int. J. Hydrogen Energy* 48, 21189–21213. https://doi.org/10.1016/j.ijhydene.2023.03.029

Sharma, P., Sahoo, B.B., Said, Z., Hadiyanto, H., Nguyen, X.P., Nižetić, S., Huang, Z., Hoang, A.T., Li, C., 2023c. Application of machine learning and Box-Behnken design in optimizing engine characteristics operated with a dual-fuel mode of algal biodiesel and waste-derived biogas. *Int. J. Hydrogen Energy* 48, 6738–6760. https://doi.org/10.1016/j.ijhydene.2022.04.152

Sharma, P., Said, Z., Kumar, A., Nižetić, S., Pandey, A., Hoang, A.T., Huang, Z., Afzal, A., Li, C., Le, A.T., Nguyen, X.P., Tran, V.D., 2022. Recent Advances in Machine Learning Research for Nanofluid-Based Heat Transfer in Renewable Energy System. *Energy & Fuels* 36, 6626–6658. https://doi.org/10.1021/acs.energyfuels.2c01006

Sharmila, V.G., Shanmugavel, S.P., Banu, J.R., 2024. A review on emerging technologies and machine learning approaches for sustainable production of biofuel from biomass waste. *Biomass and Bioenergy* 180, 106997. https://doi.org/10.1016/j.biombioe.2023.106997

Shi, J., Lee, W.-J., Liu, Y., Yang, Y., Wang, P., 2012. Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines. *IEEE Trans. Ind. Appl.* 48, 1064–

1069. https://doi.org/10.1109/TIA.2012.2190816
Simão, M.L., Videiro, P.M., Silva, P.B.A., de Freitas Assad, L.P., Sagrilo, L.V.S., 2020. Application of Taylor diagram in the evaluation of joint environmental distributions' performances. *Mar. Syst. Ocean Technol.* 15, 151–159. https://doi.org/10.1007/s40868-020-00081-5
Tanveer, M., Rajani, T., Rastogi, R., Shao, Y.H., Ganaie, M.A., 2022. Comprehensive review on twin support vector machines. *Ann. Oper. Res.* https://doi.org/10.1007/s10479-022-04575-w
Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* 106, 7183–7192.
The-Thanh, L., Tien-Long, B., The-Van, T., Duc-Toan, N., 2019. A study on a deep-drawing process with two shaping states for a fuel-filter cup using combined simulation and experiment. *Adv. Mech. Eng.* 11, 1–11. https://doi.org/10.1177/1687814019872674
Tira, H.S., Tsolakis, A., Turner, D., Herreros, J.M., Dearn, K.D., Theinnoi, K., Wyszynski, M.L., 2014. Influence of Fuel Properties, Hydrogen, and Reformate Additions on Diesel-Biogas Dual-Fueled Engine. *J. Energy Eng.* 140. https://doi.org/10.1061/(ASCE)EY.1943-7897.0000173
Tracking SDG 7 – The Energy Progress Report 2022 [WWW Document], n.d.
Usman, M., Hussain, H., Riaz, F., Irshad, M., Bashir, R., Shah, M.H., Zafar, A.A., Bashir, U., Kalam, M.A., Mujtaba, M.A., Soudagar, M.E.M., 2021. Artificial Neural Network Led Optimization of Oxyhydrogen Hybridized Diesel Operated Engine. *Sustain.* 13, 9373. https://doi.org/10.3390/SU13169373
Vasan, V., Sridharan, N.V., Feroskhan, M., Vaithiyanathan, S., Subramanian, B., Tsai, P.C., Lin, Y.C., Lay, C.H., Wang, C.T., Ponnusamy, V.K., 2024. Biogas production and its utilization in internal combustion engines - A review. *Process Saf. Environ. Prot.* 186, 518–539. https://doi.org/10.1016/J.PSEP.2024.04.014
Verma, S., Das, L.M., Bhatti, S.S., Kaushik, S.C., 2017. A comparative exergetic performance and emission analysis of pilot diesel dual-fuel engine with biogas, CNG and hydrogen as main fuels. *Energy Convers. Manag.* 151, 764–777. https://doi.org/10.1016/j.enconman.2017.09.035
Veza, I., Afzal, A., Mujtaba, M.A., Tuan Hoang, A., Balasubramanian, D., Sekar, M., Fattah, I.M.R., Soudagar, M.E.M., EL-Seesy, A.I., Djamari, D.W., Hananto, A.L., Putra, N.R., Tamaldin, N., 2022a. Review of artificial neural networks for gasoline, diesel and homogeneous charge compression ignition engine. *Alexandria Eng. J.* 61, 8363–8391. https://doi.org/10.1016/j.aej.2022.01.072
Veza, I., Karaoglan, A.D., Ileri, E., Kaulani, S.A., Tamaldin, N., Latiff, Z.A., Muhamad Said, M.F., Hoang, A.T., Yatish, K.V., Idris, M., 2022b. Grasshopper optimization algorithm for diesel engine fuelled with ethanol-biodiesel-diesel blends. *Case Stud. Therm. Eng.* 31, 101817. https://doi.org/10.1016/j.csite.2022.101817
Walker, A.M., Cliff, A., Romero, J., Shah, M.B., Jones, P., Felipe Machado Gazolla, J.G., Jacobson, D.A., Kainer, D., 2022. Evaluating the performance of random forest and iterative random forest based methods when applied to gene expression data. *Comput. Struct. Biotechnol. J.* 20, 3372–3386. https://doi.org/10.1016/j.csbj.2022.06.037
Yu, F.W., Ho, W.T., Wong, C.F.J., 2024. Predicting and decarbonizing carbon emissions from building energy use in Hong Kong: A LASSO regression approach. *Energy Sustain. Dev.* 78, 101374. https://doi.org/10.1016/J.ESD.2023.101374
Zhao, R., Xu, L., Su, X., Feng, S., Li, C., Tan, Q., Wang, Z., 2020. A Numerical and Experimental Study of Marine Hydrogen–Natural Gas–Diesel Tri–Fuel Engines. *Polish Marit. Res.* 27, 80–90. https://doi.org/10.2478/pomr-2020-0068