

# Prediction and optimization of performance and emission characteristics of a dual fuel engine using machine learning

Krishnasamy Karunamurthy<sup>1</sup>, Mohammed Musthafa Feroskhan<sup>1</sup>, Ganesan Suganya<sup>2,\*</sup>, and Ismail Saleel<sup>3</sup>

<sup>1</sup> School of Mechanical Engineering, Vellore Institute of Technology (VIT) Chennai, Tamilnadu, India

<sup>2</sup> School of Computer Science and Engineering, Vellore Institute of Technology (VIT) Chennai, Tamilnadu, India

<sup>3</sup> Department of Mechanical Engineering, National Institute of Technology (NIT) Calicut, Kerala, India

Received: 27 January 2021 / Accepted: 3 February 2022

**Abstract.** The current research in engine, fuel and lubricant development are aiming towards environmental protection by reducing the harmful emissions. The testing under various conditions becomes mandatory before releasing product to meet the sustainable development goals of United Nations. This experimentation and testing under various operating conditions is time-consuming and tiresome process; it also leads to wastage of manpower, money, precious time and scarce resources. Intelligent techniques like Machine Learning (ML) has proven it's usage in almost all domains, trying to simulate the results as trained. This advantage is used to predict the performance and emission characteristics of a dual fuel engine. In this study, the experimental data are obtained from a single cylinder CI engine by operating under dual fuel mode using biogas and diesel as primary and secondary fuel respectively. The input parameters such as biogas flow rate, methane fraction (MF), torque and intake temperature are considered to predict the output parameters. The output parameters of the study includes performance attributes Brake thermal efficiency, secondary fuel energy ratio, and emissions attributes HC, CO, NOx and smoke. The proposed model uses Random forest Regressor and is trained using 324 distinct experiences recorded through physical experimentation. The model is validated using  $R^2$  score which is observed to be 0.997 for the given dataset while trained and tested in the ratio of 85:15. The outputs of the model are used to compute the output data for any new values of input attributes. The optimized values of the input parameters that could give maximum thermal efficiency and minimum emission is found using Lagrangian optimization. The optimized values are 12.48 Nm torque, 8.29 lit/min of biogas flow rate, methane fraction of 72.8%, intake temperature of 68.3°C.

**Keywords:** Biogas / dual fuel / cross-fold validation / random forest regressor / Lagrangian optimization

## 1 Introduction

The constant depletion of conventional fuels and stringent emission norms are motivating researchers to find better alternative fuel for stationary and automotive engines. Biogas is one of the promising alternative fuels for IC engines. It is a mixture of methane (50–70%), carbon dioxide (30–45%) and other gases. Presence of carbon dioxide reduces calorific value and ignitability of biogas. Increase in methane fraction improves the calorific value of biogas. Removal of carbon dioxide ensures high methane fraction and it can be done through various processes such as water scrubber, chemical separation, membrane separation, pressure swing separation and cryogenic separation methods. It is very difficult to ignite biogas without pilot fuel due to high self-ignition

temperature of biogas. Increasing the intake temperature may enhance the combustion of biogas. Therefore, it is significant for the researchers to model and quantify the association between input characteristics and the different gaseous that are emitted. Finding optimal values for input parameters of engine to get minimal quantity of exhaust gases with increased efficiency becomes vital for justification of usage. Huge quantum of experimentation is required to find the optimal values which in real time consumes huge resources including fuels, human resource, budget, time etc.,

Industry 4.0 focuses on digital transformation in industries requiring robotization of most production processes thereby reducing resource utilization at all levels. In alignment with Industry 4.0, there has been lot of works carried out to develop predictive models using machine learning approaches thereby reducing manpower, fuel, time, and budget. The major contributions of the paper are given below.

\* e-mail: [suganyakaruna@gmail.com](mailto:suganyakaruna@gmail.com)

The data obtained from experimentation including biogas flow rate, methane fraction, torque and intake charge temperature as input parameters for engine and various gaseous emissions as output parameters are used to develop a predictive model using Random Forest Regressor. The model is validated using various metrics and with testing data. Finally, the optimized values of input parameters that may result in minimized gaseous emission is computed using Lagrangian optimization.

The paper is organized as follows. A detailed literature review on biofuel usage and utilization of predictive approaches in thermal industry is discussed in Section 2. The experimental set up and methodology is discussed in Section 3. Modeling using data collected from experiments is explained in Section 4. A detailed analysis with comparison towards predicted and experimental data is detailed in Section 5. Optimization procedure to arrive at probable input values is explored in Section 6. Finally, conclusions are emphasized and presented in Section 7.

## 2 Literature review

Two possible methods are available to utilize biogas in CI engine such as HCCI (Homogeneous Charge Compression Ignition) and dual fuel mode. In dual fuel mode, pilot fuel is used along with biogas which is having high self-ignition temperature. Addition of biogas in CI engines reduces brake thermal efficiency due to the presence of CO<sub>2</sub> in biogas. CO<sub>2</sub> reduces flame speed and combustion temperature [1,2]. Combustion temperature can be increased using enhanced compression ratio which improves combustion and reduces HC and CO emissions [3]. Up to 48% diesel substitution was achieved by Duc and Wattanavichien [4] observed around 50% diesel substitution for dual fuel mode in IDI (Indirect Injection) engine.

Fuel conversion efficiencies are low for dual fuel mode compared to diesel-only mode at low and medium loads. However, same fuel conversion efficiencies were observed at high load conditions. Comparable BSEC (Brake Specific Energy Consumption) has also been stated by Mustafi et al. [5] for the two modes. Varying injection amounts of the pilot fuel will regulate the emission parameters [5,6]. The brake thermal efficiency can be improved by an increase in the intake temperature, compression ratio and oxygen content [3,7,8]. Supplying extra oxygen decreases the volatility of combustion and increases diesel replacement. This reduces the methane emitted, while the effects on CO emissions is inconsistent [7]. In dual fuel mode, Sorathia and Yadav [9] noted that coolant losses were higher and exhaust losses lower, the net result being thermal efficiency similar to that of diesel-only mode. In dual fuel mode, reduced percentage exergy destruction and greater exergy efficiency were recorded.

Barik and Murugan [10] have studied biogas fueled dual fuel operation using 0–0.6 kg/h biogas flow rate. The replacement of up to 30% diesel was accomplished at full loading. Due to CO<sub>2</sub> displacing more air and decreasing the burning rate, volumetric efficiency declined and BSEC increased to high values in dual fuel mode. Better performance and emission indices are observed at 0.9 kg/h

biogas flow rate. Rahman and Ramesh [11] examined methane fraction effects in the 24–68% range. The impact of the full-range methane fraction (from raw biogas to pure methane) is, however, not investigated. The presence of CO<sub>2</sub> in biogas has an impact comparable to EGR, minimizing NO<sub>x</sub>, but increasing emissions of CO and HC. Homogeneous mixture of biogas replaces heterogeneous nature of diesel and reduces smoke emissions [12]. Barik and Sivalingam [13] recorded low exhaust temperature, high HC and CO emissions and low NO<sub>x</sub> and smoke emissions.

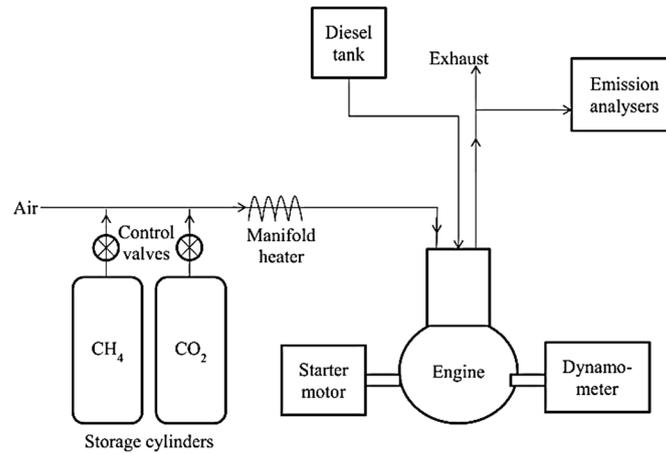
Utilization of prediction models in mechanical industry is discussed by many researchers. Girish Kanta et al. [14] predicted the machining parameters using Artificial Neural Networks coupled with Genetic algorithm and has proved the agreement of predicted data with experimental data through validation. Pham et al. [15] discussed about the usage of random forest combined with particle swarm optimization to predict undrained shear strength of soil.

## 3 Experimental setup

In this investigation, a regular single-cylinder 4-stroke direct injection CI engine (6 kW maximum power) was used. An arrangement for the introduction of simulated biogas for dual fuel operation was provided for the test setup (see Fig. 1). By blending compressed CH<sub>4</sub> and CO<sub>2</sub> delivered from separate cylinders through pressure controlling valves and thermal mass flow meters, biogas was synthesized. In order to adjust the composition of the simulated biogas, the gas flow rates were regulated independently. Biogas purity was expressed in terms of the concentration of methane (by volume). To put in the CH<sub>4</sub> and CO<sub>2</sub>, a Y-shaped nozzle was connected to the manifold, which was eventually combined in a cylindrical chamber. To ensure that the piston and valve-induced flow oscillations did not affect the flow meter readings, the nozzle was located sufficiently upstream of the cylinder. Further downstream, to preheat the biogas-air mixture, a honeycomb structured heater was mounted in the intake manifold. Using a voltage regulator, the heating rate was controlled. To identify air flow rate and fuel flow rate, an orifice meter and a burette were used respectively. Using a portable gas analyser and smoke meter, the four main emissions of exhaust gases (HC, CO, NO<sub>x</sub> and smoke) were found.

### 3.1 Experimental methodology

A detailed factorial analysis was conducted to investigate the effects on performance and emission characteristics of biogas flow rate, methane fraction, torque and intake charge temperature under biogas-diesel dual fuel operation. The range of each input parameter was chosen to cover the engine's entire operating range. The biogas flow rate range (2–16 l/min) is equal to 10–90% of the overall energy released during combustion. Methane fraction amounts, viz. The availability of methane in naturally derived biogas and pure methane is 50–100%. In order to



**Fig. 1.** Schematic diagram of the experimental setup.

**Table 1.** Operating parameters.

Levels	Applied load (Nm)	Biogas flow rate, $Q_{\text{biogas}}$ (lpm)	Methane fraction (%)	Intake temperature, $T_{\text{intake}}$ (°C)
Level 1	5	2	50	35
Level 2	10	4	65	60
Level 3	15	8	80	80
Level 4	20	12	100	100
Level 5	–	16	–	–

investigate the effects of charge preheating, the intake temperature was configured to differ from room conditions (35 °C) to the actual working limit of the heater (100 °C). At a constant rate of 1900 rpm, the applied load ranged from 20 to 90% of the rated engine load. Operating parameters are provided in Table 1. The method offered by Moffat [14] is used to find uncertainties in the output parameters. For all parameters, less than 3% uncertainties were noticed. A total of 324 independent observations are done and is stored in an excel file. The combination of operating parameters along with emission characteristics is shown in Table 2.

## 4 Modeling of system using machine learning

Data collected from physical experimentation is stored using Microsoft excel. Load, Biogas flowrate, Methane fraction and intake temperature are considered to be predictor variables and Carbon-monoxide, Hydrocarbon, smoke, Nitrogen oxide and brake thermal efficiency are considered to be response variables. Around 325 independent experimentations are done by varying input parameters and the results are recorded. Initial analysis of data is presented in Table 3.

The data is preprocessed, trained using random forest and the results are used for finding optimal values of input features. Figure 2 depicts the steps carried out in arriving at the optimal value.

### 4.1 Data preprocessing

Every value received from experimentation is recorded manually and hence are prone to errors. Before proceeding to modeling, the entire dataset is preprocessed to align the values suitable for processing by machine learning algorithms. The following activities are done:

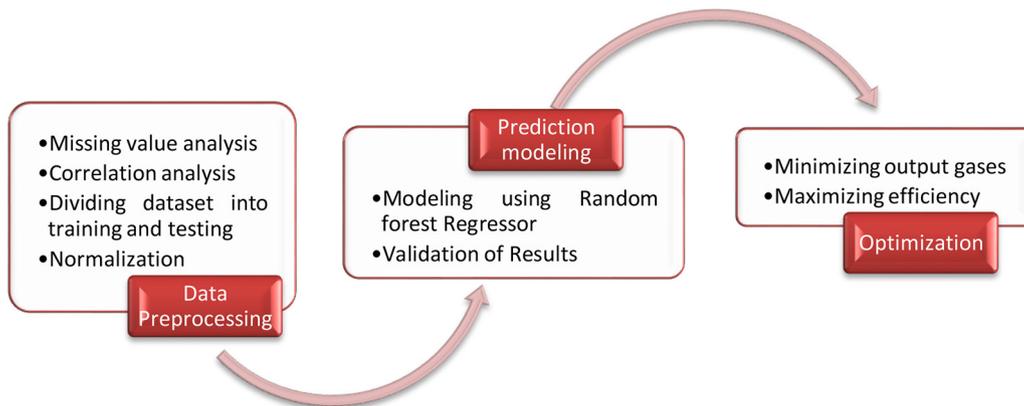
- The dataset is checked for missing values. Two features (Load, Intake temperature) are found to have 15 data missing in common. Mean of each feature is calculated and are used to fill up the missing values.
- Correlation analysis is then performed to understand the linearity between predictor and response variables. In particular, correlation between input variables and each output variable is calculated using Pearson's correlation coefficient. Values or correlation range from  $-1$  to  $+1$  indicating positive to negative correlation. Linearity exists if the values are either near to  $+1$  or  $-1$ . Figures 3 and 4 presents the correlation variables. The correlation between the input and output features are very low lying in the midst between  $-1$  and  $+1$ , thus justifying the usage of methods suitable for non-linear models.
- The dataset, D is divided into training and testing sets in the proportion of 80:20. Recommended ratio of training and testing samples 70:30, 75:25, 80:20, 85:15, 90:10 are used and sensitivity analysis is done. The division is randomized to reduce bias towards selection of instances.

**Table 2.** Recorded values of operating parameters with emission.

Applied load (Nm)	Biogas flow rate, $Q_{\text{biogas}}$ (lpm)	Methane fraction (%)	Intake temperature, $T_{\text{intake}}$ ( $^{\circ}\text{C}$ )	Carbon monoxide (%)	Hydro carbon (ppm)	Smoke (%)	Nitrogen oxide (ppm)	Brake thermal efficiency (%)
5	2	50	35	0.13	105	35	96	18.50771
5	2	65	35	0.12	143	33	104	18.41667
5	2	80	35	0.13	167	33	118	18.54554
5	2	100	35	0.15	189	31	103	18.38235
10	2	50	35	0.12	107	41	267	28.59791
10	2	65	35	0.11	123	36	257	28.60493
10	2	80	35	0.12	146	41	262	28.37363
10	2	100	35	0.12	155	40	253	27.98274
15	2	50	35	0.09	94	44	607	32.11983
15	2	65	35	0.09	109	43	605	32.47331
15	2	80	35	0.1	116	42	546	32.80313
15	2	100	35	0.09	133	41	589	31.74158
20	2	50	35	0.19	74	54	755	33.58341
20	2	65	35	0.19	76	52	805	33.79799
20	2	80	35	0.18	104	51	870	33.67402

**Table 3.** Analysis of dataset.

	Load	Biogas flow rate	methane fraction	Intake temperature	CO	HC	Smoke	Nox	BT (%)
Count	324	324	324	324	324	324	324	324	324
Mean	12.48765	8.296296	72.83951	68.33333	0.167099	256.7901	31.97346	430.3673	26.39792
Std	5.61647	5.182596	20.13789	24.25356	0.058619	149.8792	10.54105	340.2342	7.173296
Min	4	0	0	35	0.07	51	10	10	8.392
25%	8.75	4	50	35	0.12	140.75	24	133.25	20.16379
50%	12.5	8	65	60	0.17	216.5	31	334.5	29.29777
75%	16.25	12	80	80	0.2	346.25	41	712.75	32.25244
Max	20	16	100	100	0.38	793	57.6	1260	35.06525



**Fig. 2.** Modeling methodology.

	CO	HC	Smoke	Nox	BT (%)
Load	0.395975	-0.447559	0.744710	0.868802	0.918328
Biogas flow rate	0.631416	0.757641	-0.586702	-0.317312	-0.180979
methane fraction	0.109286	0.328115	-0.221814	0.011166	-0.048128
Intake temperature	-0.015715	-0.039491	-0.078573	0.225935	-0.016137

Fig. 3. Correlation between input and output variables.



Fig. 4. Heatmap representing correlation.

– The values are recorded in different scales and each feature may have different numeric range. Variations in the scales across input variables may increase the trouble in understanding the results of the problem being modeled. Since, algorithms tend to work with numeric values; each data is considered quantitatively and hence will create an impact on the modeling. When the ranges of each feature are different, the impact on the results will be large. Hence, to avoid such anomaly, data is normalized to a range of 0 and 1 using min–max normalization. Training dataset is normalized to reduce anomaly due to varying numeric ranges. Figure 5 presents the snapshot of normalized values corresponding to input parameters.

[	0.25	0.5	0.69230769	0.12903226]
[	0.25	0.65	0.69230769	0.09677419]
[	0.25	0.8	0.69230769	0.09677419]
[	0.25	1.	0.69230769	0.12903226]
[	0.25	0.5	0.69230769	0.06451613]

Fig. 5. Normalized values of predictor variables.

## 4.2 Prediction modeling

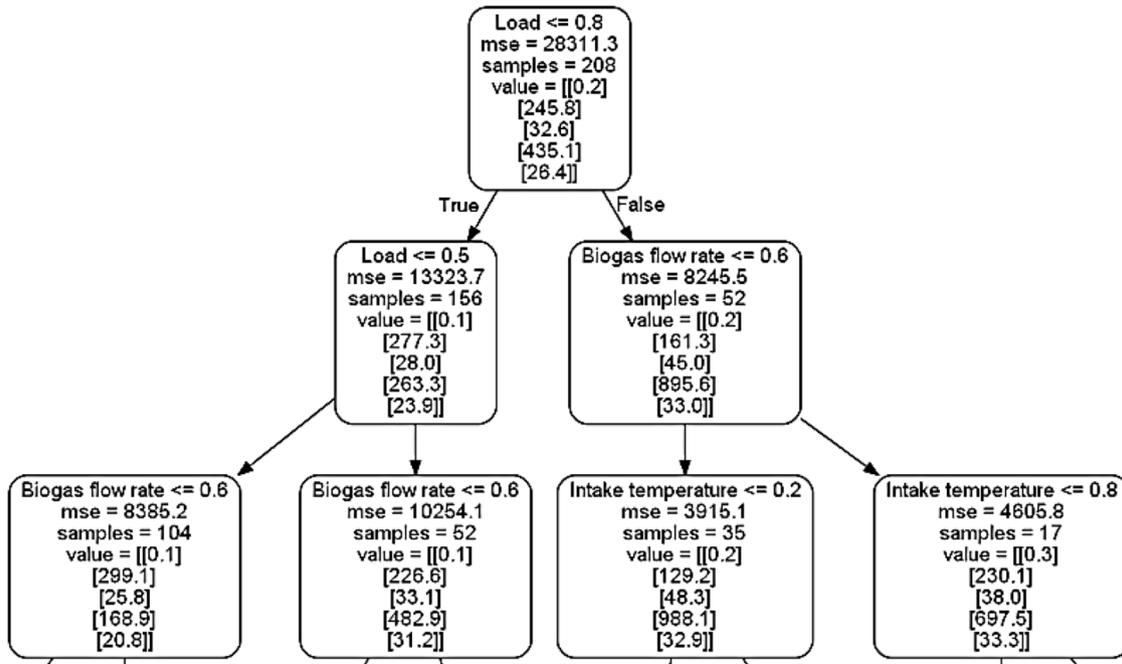
### 4.2.1 Random forest regressor

Random forest is a supervised ensemble learning algorithm that could be used both for regression and classification. Based on Correlation and Regression Trees (CART), this algorithm constructs a swarm of decision trees during training time. The regression trees are created through a process called bootstrapping, a process of drawing random

samples from the dataset with replacement. Constructed regression trees are then used as base learners. Each node in a regression tree represents a binary test against the selected predictor variable. The selected variable is used to minimize the MSE (residual sum of squares) for the data flowing down left and right branches. Mean of all the predicted values from each tree is then considered to be the final prediction. The advantage of CART is that it can fit the data very well and may result in low bias. But, since the results completely depend on input data, CART may suffer from high variance. This can be adjusted by selecting ‘ $m$ ’ randomly selected predictors,  $m \leq n$  (‘ $n$ ’ represents the actual number of predictors) for construction each tree and by combining the results and hence Random forest is proven to reduce high variance and overfitting problems. The performance may be validated using  $R^2$  as the metric.

**Table 4.** Hyperparameter details.

SI. No	Hyperparameter	Definition	Range varied / Value set
1	No_estimators	Number of trees that could be formed	5 to 85 in steps of 5
2	Max_depth	Maximum depth of the tree	3 to 15 in step of 1
3	Min_samples_leaf	Minimum number of samples required to be the leaf node	3 to 1
4	Min_samples_split	Minimum number of samples to split a node	3 to 1



**Fig. 6.** Random Forest sample for training dataset.

4.2.2 Algorithm

Input: Dataset, D of size 324 × 9 with values stored for predictor variables [Load, Biogas flow rate, Methane fraction, Intake temperature] and response variables [Carbon-monoxide, Hydrocarbon, Smoke, Nitrogen Oxide, Brake Thermal Efficiency].

Output: Predicted values for response variables based on inputs for predictor variables.

Method:

- Create a random forest using the following steps:
- Create a bootstrap sample D' from the dataset D.
- Construct a decision tree for the bootstrapped sample.
- Steps (a) and (b) are repeated for 'n' number of times, where 'n' is the number of subtrees required or possible.
- Calculate predictive value from each tree. The mean of the predicted values represent the output.

4.2.3 Implementation

Random forest regressor is used to model the nonlinear relationship from predictor variables to response variables. This is implemented using Python with Jupyter notebook as frontend. The random forest model is trained using the normalized training dataset. Optimal values of response

variables are found by tuning various hyperparameters. Table 4 presents the various hyperparameters used for tuning the performance of the model.

With fine tuning, a regression forest is built part of which is shown in Figure 6.

Deviation observed between expected and actual value is termed as error or loss. Model evaluation is done using Root Mean Square Error (RMSE) as given in equation (1). Figure 7 represents the actual and predicted values of all response variables (Carbon Monoxide, Hydrocarbon, Smoke, Nitrous Oxide and Brake thermal efficiency). Values vary from 0.07 (minimum value of carbon monoxide) to 1260 (maximum value of Nitrous oxide).

$$RMSE = \sqrt{\frac{1}{no} \sum_{i=1}^{no} (value_{actual} - value_{predicted})^2} \tag{1}$$

4.2.4 Optimization

The use of optimization in various applications is discussed by researchers [16,17]. Once the modeling is done, the final task is to predict the values of input features that could

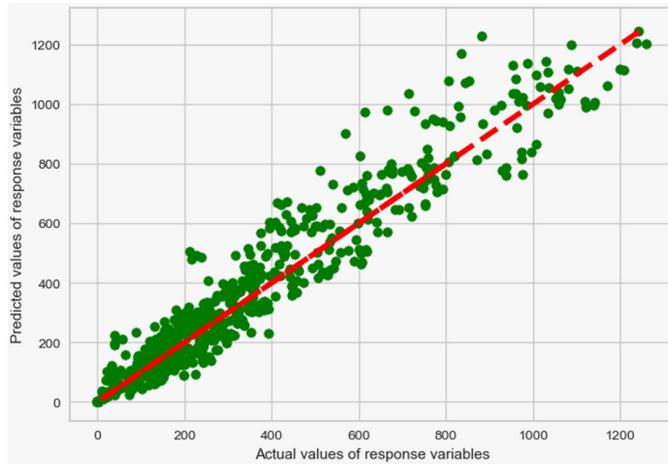


Fig. 7. Actual vs predicted values of response variables.

Table 5. Range of input variables.

Input Variable	Range
Load	4–20
Biogas flow rate	0–16.0
Methane Fraction	0–100.0
Intake temperature	35.0–100.0

produce minimum exhaust gases with increase in brake thermal efficiency. This is performed by solving constrained optimization using Lagrangian optimizer. The boundary values for the input variables are shown in Table 5.

Two optimizations are performed, first to minimize the exhaust gas output and second to maximize the brake thermal efficiency. The boundary conditions as stated in Table 5 are used as constraints for both optimizations. The output of objective function is computed using the modeling created using Random forest Regressor. The results of two optimizations are averaged to arrive at the final results.

## 5 Results and discussion

The  $R^2$  values are recorded by varying different proportions of training and testing data. Figure 8 showing the  $R^2$  for different proportions of training and testing data reveals that the model is stabilized above 85:15 ratios. The score towards 100% reveals that the model could represent all variability of output data.

The proposed model is compared with  $R^2$  score of models trained with multilinear regression, support vector regression and KNN algorithms. Figure 9 represents the comparisons of models based on  $R^2$  score. The proposed model outperforms other model and the results are satisfactory.

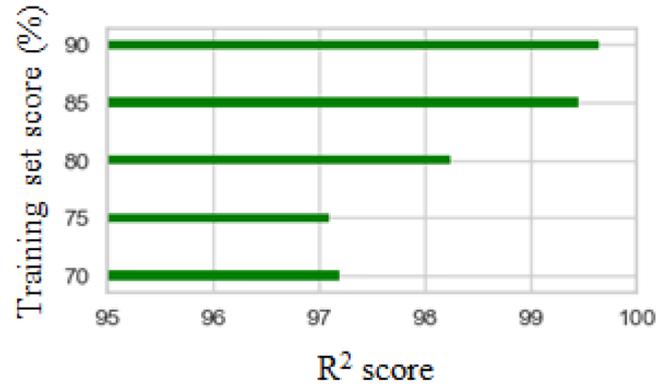


Fig. 8.  $R^2$  score for different proportions of testing and training data.

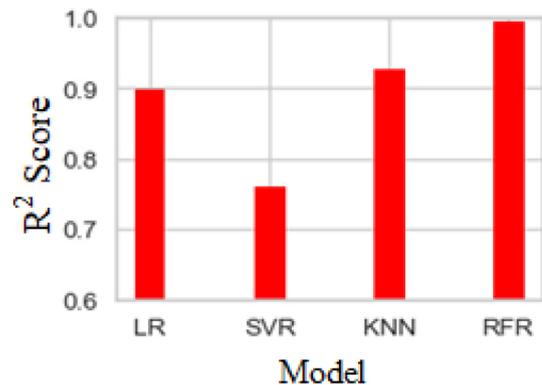


Fig. 9. Comparison of  $R^2$  score for various models.

## 6 Conclusions

The environmental degradation can be reduced and prevented by optimizing the dual fuel engine operating parameters by following machine learning techniques for prediction and lagrangian optimization. The optimized operating parameters such as torque, bio gas flow rate, methane fraction and fuel intake temperature for minimum emission and maximum thermal efficiency are 12.48 Nm, 8.29 l/min, methane fraction of 72.8%, intake temperature of 68.3°C respectively.

## References

- S.S. Kalsi, K.A. Subramanian, Effect of simulated biogas on performance, combustion and emissions characteristics of a bio-diesel fueled diesel engine, *Renew. Energy* **106**, 78–90 (2017)
- B.K. Debnath, B.J. Bora, N. Sahoo, U.K. Saha, Influence of emulsified palm biodiesel as pilot fuel in a biogas run dual fuel diesel engine, *J. Energy Eng.* **140**, 1–9 (2014)
- B.J. Bora, U.K. Saha, S. Chatterjee, V. Veer, Effect of compression ratio on performance, combustion and emission characteristics of a dual fuel diesel engine run on raw biogas, *Energy Convers. Manag.* **87**, 1000–1009 (2014)

4. P.M. Duc, K. Wattanavichien, Study on biogas premixed charge diesel dual fuelled engine, *Energy Convers. Manag.* **48**, 2286–2308 (2007)
5. N.N. Mustafi, R.R. Raine, S. Verhelst, Combustion and emissions characteristics of a dual fuel engine operated on alternative gaseous fuels, *Fuel* **109**, 669–678 (2013)
6. A.C. Polk, C.M. Gibson, N.T. Shoemaker, K.K. Srinivasan, S.R. Krishnan, Analysis of ignition behavior in a turbo-charged direct injection dual fuel engine using propane and methane as primary fuels, *J. Energy Resour. Technol.* **135**, 2202–2212 (2013)
7. K. Cacua, A. Amell, F. Cadavid, Effects of oxygen enriched air on the operation and performance of a diesel-biogas dual fuel engine, *Biomass Bioenergy* **45**, 159–167 (2012)
8. M. Feroskhan, S. Ismail, M.G. Reddy, A. Sai Teja, Effects of charge preheating on the performance of a biogas-diesel dual fuel CI engine, *Eng. Sci. Technol.* **21**, 330–337 (2018)
9. S. Harilal, H.J.Y. Sorathia, Energy analyses to a Ci-engine using diesel and bio-gas dual fuel – a review study, *Int. J. Adv. Eng. Res. Stud.* **1**, 212–217 (2012)
10. D. Barik, S. Murugan, Investigation on combustion performance and emission characteristics of a DI (direct injection) diesel engine fueled with biogas-diesel in dual fuel mode, *Energy* **72**, 760–771 (2014)
11. K.A. Rahman, A. Ramesh, Studies on the effects of methane fraction and injection strategies in a biogas diesel common rail dual fuel engine, *Fuel* **236**, 147–165 (2019)
12. S. Swami Nathan, J.M. Mallikrajuna, A. Ramesh, Homogeneous charge compression ignition versus dual fuelling for utilizing biogas in compression ignition engines, *Proc. Inst. Mech. Eng. Part D* **223**, 413–422 (2009)
13. D. Barik, M. Sivalingam, Performance and Emission Characteristics of a Biogas Fueled DI Diesel Engine, SAE Technical Paper (2013) <https://doi.org/10.4271/2013-01-2507>
14. G. Kanta, K. Singh Sangwanb, Predictive modelling and optimization of machining parameters to minimize surface roughness using artificial neural network coupled with genetic algorithm, in *15th CIRP Conference on Modelling of Machining Operations* **31**, 453–458 (2015)
15. B.T. Pham, C. Qi, L.S. Ho, T. Nguyen-Thoi, N. Al-Ansari, M.D. Nguyen, H.D. Nguyen, H.-B. Ly, H.V. Le, I. Prakash, A novel hybrid soft computing model using random forest and particle swarm optimization for estimation of undrained shear strength of soil, *Sustainability* **12**, 2218 (2020)
16. A. Mosavi, Application of data mining in multiobjective optimization problems, *Int. J. Simul. Multisci. Des. Optim.* **5**, A15 (2014)
17. R. Madhu Kumar, N.V.V.S. Sudheer, K. Ganesh Babu, Multi-attribute decision making parametric optimization in two-stage hot cascade vortex tube through grey relational analysis, *Int. J. Simul. Multidisci. Des. Optim.* **11**, 21 (2020)

**Cite this article as:** Krishnasamy Karunamurthy, Mohammed Musthafa Feroskhan, Ganesan Suganya, Ismail Saleel, Prediction and optimization of performance and emission characteristics of a dual fuel engine using machine learning, *Int. J. Simul. Multidisci. Des. Optim.* **13**, 13 (2022)