Journal of Engineering and Technological Sciences

Research Paper

Forecasting of Engine Performance for Gasoline-Ethanol Blends using Machine Learning

Shailesh Sonawane[1,2], Ravi Sekhar[3,*], Arundhati Warke[3], Sukrut Thipse[2] & Chetan Varma[2]

[1]Research Scholar, Symbiosis Institute of Technology (SIT), Pune Campus, Symbiosis International (Deemed University) (SIU), Pune, 412115, Maharashtra, India
[2]Automotive Research Association of India (ARAI), Pune, Maharashtra, India
[3]Symbiosis Institute of Technology (SIT), Pune Campus, Symbiosis International (Deemed University) (SIU), Pune, 412115, Maharashtra, India

Corresponding author: ravi.sekhar@sitpune.edu.in

Abstract

The incorporation of alternative fuels in the automotive domain has brought a new paradigm to tackle the environmental and energy crises. Therefore, it is of interest to test and forecast engine performance with blended fuels. This paper presents an experimental study on gasoline-ethanol blends to test and forecast engine behavior due to changes in the fuel. This study employed a machine learning (ML) technique called TOPSIS to forecast the performance of a slightly higher blend fuelled engine based on experimental data obtained from the same engine running on 0% ethanol blend (E0) and E10 fuels under full load conditions. The engine performance predictions of this ML model were validated for 15% ethanol blend (E15) and further used to predict the engine performance of 20% ethanol blend fuel. The prediction R^2 score for the ML model was found to be greater than 0.95 and the MAPE range was 1% to 5% for all observed engine performance attributes. Thus, this paper presents the potential of TOPSIS methodology-based ML predictions on blended fuel engine performance to shorten the testing efforts of blended fuel engines. This methodology may help to faster incorporate higher blended fuels in the automotive sector.

Keywords: *E0; E10; engine performance; ethanol; machine learning.*

Introduction

With the advancement in technology, the world is inherently orienting itself to data with more complexity. Intuitively, this advancement captures physical processes with more and more precision. Precision in data mining of a physical process explicitly gives rise to contemporary problems that are complex in nature. This breakdown of the process needs to be well understood to approach a problem. In this regard, artificial intelligence has evolved over the years to come up with various insights and hidden patterns to help in forecasting as well as data driven decision making [1,2].

The umbrella of artificial intelligence (AI) includes various methodologies that serve a crucial purpose in crossing hurdles in contemporary problem statements. An important one is machine learning (ML). ML methods have proven to be very effective in addressing various problems in different domains, such as combustion [3,4], manufacturing [5,6], aerodynamics [7], and battery management [8]. Improvement of automotive engines has been widely investigated for over a decade when the idea of alternative fuels was first introduced in the automotive domain [9]. Alternative fuels not only serve to support vehicle motion but also mitigate the dependency on conventional sources of energy. A lot of work has been done in this domain to improve engines and address complexities like dual fuel use [10], materials compatibility [11], CNG engine development [12,13], and methanol blending [14]. These fuels include gasoline doped with ethanol and methanol; diesel doped with ethanol and methanol; diesel doped with CNG and LNG, H_2 doped with CNG, etc.

Ethanol as a fuel has a higher-octane number and oxygen content compared to gasoline, which enhances combustion and complete fuel burning. The fuel being renewable in nature can boost the agricultural sector and mitigate the dependency on fossils. India is on the verge of incorporating E20 (20% ethanol blended in gasoline

ISSN: 2337-5779
J. Eng. Technol. Sci. Vol. 55, No. 3, 2023, 340-355
DOI: 10.5614/j.eng.technol.sci.2023.55.3.10

DOI: 10.5614/j.eng.technol.sci.2023.55.3.10

by volume) on a national level by 2025. This elevated interest of incorporating ethanol as primary fuel is to utilize energy from agricultural feedstock. This benefits the overall reduction of oil imports, which will benefit the overall economy [11]. The massive consumption of non-renewable resources and environmental pollution threaten to cause an energy crisis, which will require us to create more restrictive regulations to exercise environmental scrutiny. With the introduction of the BS VI norm in India, future regulations are bound to become more stringent with standards under consideration for unregulated emissions such as aldehyde and ketone group formation. The conventional methods employing CFD models and various analytical and numerical simulation methods are complex and time consuming in forecasting engine combustion dynamics [15,16], performance [17], and emissions and engine knock [18,19]. This brings in the contemporary method of data driven modelling to aid in the forecasting of engine performance and emissions, which proffers scientific driven decision making.

Novelty of the Present Study

This study employed an ensemble learning method to train an ML model on E0 and E10 fuels, which was validated with experimental runs for E15 fuel, after which the performance data for E20 fuel were predicted to check for this approach's reliability. We also attempted to showcase the usage of multi-criteria decision-making techniques, which may be very useful along with machine learning to prepare for optimized experimental runs to predict the best engine operating points and the worst engine operating points, which may be useful for engine calibration. Based on the literature survey done by the authors, it was observed that the use of machine learning to date has limited use in engine calibration. This methodology will also help to reduce overall testing time and cost for research projects in this specific domain.

Literature Review

Conventional sources of energy are being consumed at ever higher degrees as transportation fuel, significantly contributing to environmental pollution. The prospect of a clean and green environment for future survival greatly magnifies the importance of alternative fuels for transportation. That being said, the research in this field needs thorough comprehension in regard to the performance of such fuels, emission mitigation, and the longevity of the fuel components under controlled circumstances. Chandrasekar, *et al.* [20], in their review study economically assessed and thoroughly laid down vegetable oil, bio-ethanol, glycerol and bio-diesel as some of the alternatives parried to diesel. Bielaczyc, *et al.* [21] visualized the effect of the ethanol blend percentage in gasoline for a light-duty vehicle on E5, E10, E25, and E50 on NEDC, UDC, and EDC test cycles. The study summarized the effect of properties of fuel on regulated and unregulated emissions and found that the blend E25 showed extreme responses of different emission species. The advantage of incorporating ethanol as a fuel is that it has higher auto ignition temperature, latent heat of vaporization, and flash point, making it easier for storage and transportation [22]. However, due to its lower calorific value more fuel will be consumed to extract the same amount of output power as compared to gasoline using the same engine type. The literature on this topic is quite vast and complex, but prospective alternative fuels are moving quickly towards execution strategies.

The potential of AI and ML to computationally map the process could come in very handy for the automotive industry to forecast engine performance with higher accuracy instead of employing time-consuming simulation methods based and numerical based models. Literature reviews in regard to the usage of machine learning in this domain seems to be relatively scarce but have enticed a lot of interest among researchers and the industry. Zhou, *et al.* [23] reviewed machine learning for combustion, summarizing innovative usage of machine learning for combustion instability, combustion dynamics, fuel properties prediction and design, combustion optimization, etc. This application is made difficult by the complexity of the combustion process and the ability of training data to map its chemical nature. Also considering the combustion in engines, the mapping ability of a preliminary model on the continuous operation of the engine may require recalibration and retraining [24]. Various other studies have tried to incorporate data from the literature and also experimental methods for ML algorithms such as the Gaussian process, Random Forest, Neural Networks, etc., which could be fruitful to obtain better predictions of the emissions, knock, and performance of IC engines [25-27]. In the present study, we investigated the prospect of machine learning on engine performance by conducting tests with ethanol blended

fuels, which also lays a road map for predicting engine performance when alternative fuels are variably viable on national and global levels.

Experimental Methodology

The study methodology was selected based on the idea of applying machine learning in the IC engine domain. The fundamental philosophy behind this procedure is the contemporary concept of data driven modelling and decision making. This would provide a computational complex mathematical process to explain the actual physical process and enable comparison with performance parameters in IC engines. The experimental testing procedure consisted of full throttle performance for 100% loading using two fuels, namely gasoline (E0) and ethanol-gasoline blend of 10 to 90 % by volume (E10). The fuel properties of the pure fuel and the blended fuel are given in Table 1 below.

Table 1 Properties of test fuels.

Properties	Test Method	Gasoline (E0)	Gasoline-Ethanol Blend (E10)
Density (Kg/m^3 @ 15 °C)	ASTM D 4052	749.6	752.2
Research Octane Number	ASTM D 2699	92.1	97
Calorific Value (MJ/kg)	ASTM D 4814	42.6	40.97
Reid Vapor Pressure (kPa @ 38 °C)	ASTM D 5191	53.8	55.5
Final Boiling Point (°C)	ASTM D 86	201	189.8

The full throttle performance (FTP) was performed on the engine dynamometer of a 1.2-l SI engine with four cylinders, four-stroke multi-point injection for both E0 and E10 fuels. The engine specifications are listed in Table 2 below. The experimental methodology consisted of carrying out the FTP procedure at 13 different speeds from 1,000 to 6,000 rpm with 500-rpm intervals and two other speeds, i.e., 4,400 (max torque) and 6,200 rpm (maximum power), respectively. The cycle-based data were averaged out for all the speeds and were post processed for effective data comparison.

Table 2 Test engine specifications.

Engine type	SI engine
No. of cylinders	4
Aspiration type	Naturally aspirated
Injection system	Multi-point injection
Cubic capacity (cc)	1200

Experimental Results

Experimental Data Collection

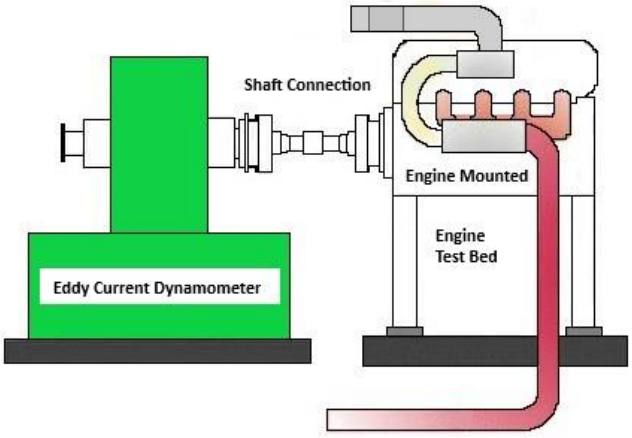
The performance of the engine was examined using the same engine type and different fuels. Data interpretation reveals the variation in performance for the same engine for various parameters. Data acquisition was performed with the engine tested on the engine dynamometer, which yielded the sets of engine performance data shown in Figure 1. The data measurement uncertainty is a stochastic process that is unavoidable in measuring devices. The engine performance measurement uncertainty for various measured parameters at 95 % confidence level is shown in Table 3.

Table 3 Measurement uncertainty for engine performance parameters.

S. No	Engine Performance parameters	Uncertainty value (@ 95 % confidence level)
1	Engine speed	± 3.5 rpm
2	Torque	± 2 N-m
3	Power	± 0.5 KW
4	Fuel flow	± 0.05 kg/hr

DOI: 10.5614/j.eng.technol.sci.2023.55.3.10

The engine was mounted on a test bed with the help of anti-vibration mounts. Then, the engine was coupled with an eddy current dynamometer with the help of a drive shaft. The eddy current dynamometer was used to apply the desired load on the engine at a certain engine speed.



Schematics of eddy current dynamometer with engine mounted for testing.

The equipment used during the data collection is listed in Table 4. To maintain uniformity in the results with different fuel blends, the engine was maintained within pre-decided boundary condition. A conditioned air system was used to provide the intake air at 25 °C ± 2 ° at 100 kPa pressure. The engine was first warmed up by applying a random load. A water temperature of 80 °C ± 2° and oil temperature of 120 °C ± 2° was achieved before starting the actual full throttle performance (FTP) test.

The first set of FTP data was taken with the E0 fuel. The engine was initially set to 6,200 rpm and 100% throttle by using the speed–throttle (N-α) mode of the dynamometer. In this typical mode, the dynamometer tries to maintain the commanded speed with 100% throttle by applying load (torque) on the flywheel. This torque applied to the flywheel is recorded as torque achieved during the set speed and throttle position. For each operating point, parameters like fuel flow, torque, speed, throttle (%), in-combustion data was captured and recorded. After completion of the first point, the engine was set to a new rpm, i.e., 6,000 rpm with 100% throttle, and again the data was recorded once the set point was stabilized. This process was repeated for new RPMs with steps of 500 rpm to 1,000 rpm.

After completion of E0 data recording, the fuel was changed to E10 and the engine was warmed up again. The same process as mentioned in the previous paragraph was repeated with the E10 fuel. The P-θ curve for all speeds were visualized and shown as a waterfall diagram. A comparison of the overall averaged out cycle-based data for E0 and E10 testing is given in the subsequent sections.

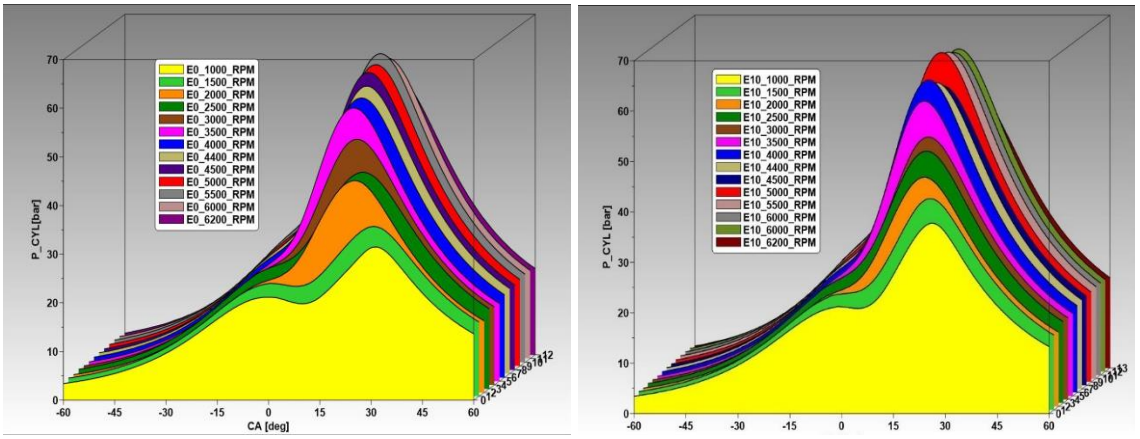
Table 4 List of equipment used during the experimentation.

S. No	Equipment	Make
1	Engine dynamometer	SAJ AG 150
2	Conditioned air system	KS, COND AIR SYSTEM - 03
3	Air flow meter	ABB SENSYFLOW – SFI - 09
4	Fuel flow meter	FEV

Engine Performance Output Comparison

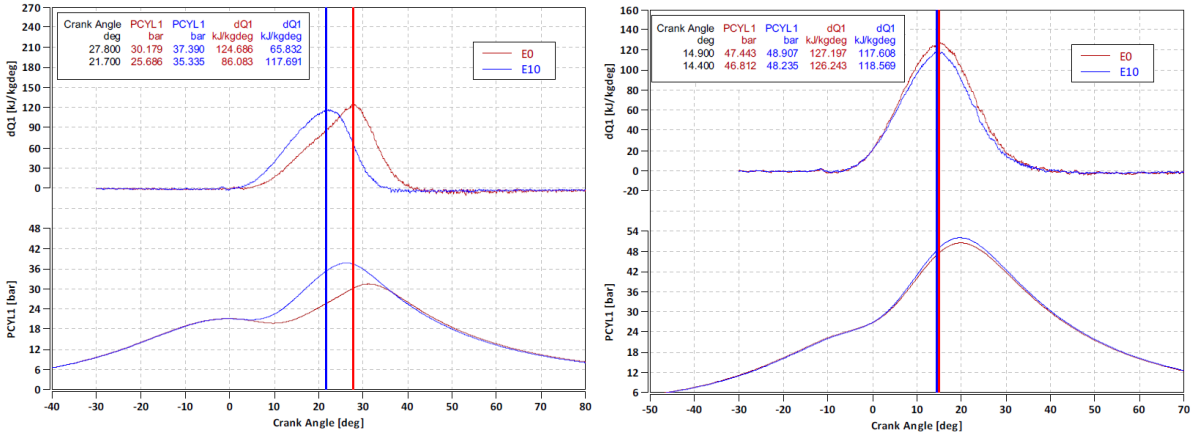
Cylinder Pressure and HRR

The comparative averaged out cylinder pressure values at all speeds were compared for the E0 and E10 fuels, so as to get visual insight during engine strokes, as shown in Figure 2. At lower-end RPM, the cylinder pressure was lower for E0 compared to E10, as indicated by the variation of the combustion duration at these points. This perhaps could be due to the higher research octane number (RON) of the fuel and the higher laminar flame speeds of the fuel, in accordance with the literature [Baddu, *et al.*, 28], which depicts better combustion at these points for E10 as compared to E0.



Waterfall diagram for P-θ of the averaged-out data for E0 and E10 at all speeds.

The averaged-out data for HRR and cylinder pressure were also processed for low, medium, and high speeds, as shown in Figures 3 and 4, respectively. This comparison also aids the improvement of heat to work conversion for E10 compared to E0, as it is an oxygenated fuel, which inherently improves the efficiency as well.



Heat release rate comparison for E0 and E10 at low and medium speed.

Table 5 shows a comparison of the combustion pressure and heat release for the E0 and E10 fuels at low speed based on the graph shown in Figure 3. In this typical low-speed graph, the blue line corresponds to the E10 fuel, whereas the red line corresponds to the E0 fuel. Two vertical lines were placed in the graph based on the highest heat release rate point for each individual fuel.

Table 5 Combustion pressure and heat release comparison at low speed.

Low Speed Category	Crank Angle (Degree)	Pcyl (bar) E0 (Red line)	Pcyl (bar) E10 (Blue line)	dQ_1 (KJ/kg deg) E0 (Red line)	dQ_1 (KJ/kg deg) E10 (Red line)
Fuel - E0	27.8	30.179	37.390	**124.686**	65.832
Fuel - E10	21.7	25.686	35.335	86.083	**117.691**

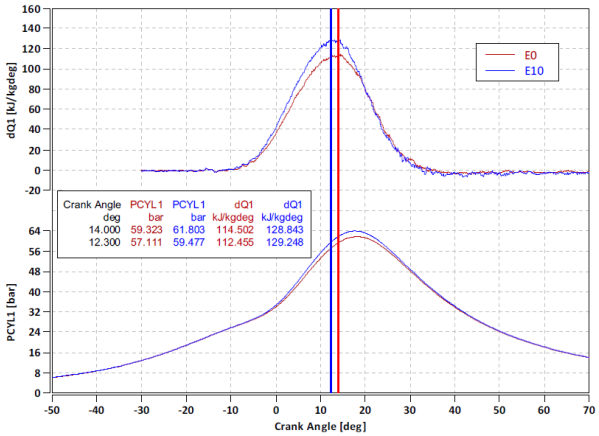
The heat release rate for E0 and E10 fuel was found to be **124.686** and **117.691** KJ/kg deg. at crank angle positions of 27.8° and 21.7° respectively. It can be seen that at lower speed, the HRR was higher for E0 when compared to E10, however it can be observed that the maximum heat was released much earlier for E10 than for E0, which means that the rate of combustion is better for the E10 fuel. Due to the early release of heat, more productive work can be generated as the produced energy can be used to push the crankshaft to a downward position and heat is not lost during the expansion stroke. This phenomenon is useful for improving the efficiency of the engine when using E10 fuel. Table 6 shows a comparison of the combustion pressure and heat release for the E0 and E10 fuels at medium speed based on the graph shown in Figure 3.

DOI: 10.5614/j.eng.technol.sci.2023.55.3.10

Table 6 Combustion pressure and heat release comparison at medium speed.

Medium Speed Category	Crank Angle (Degree)	Pcyl (bar) E0 (Red line)	Pcyl (bar) E10 (Blue line)	dQ_1 (KJ/kg deg) E0 (Red line)	dQ_1 (KJ/kg deg) E10 (Blue line)
Fuel - E0	14.9	47.443	48.907	**127.197**	117.608
Fuel - E10	14.4	46.812	48.235	126.243	**118.569**

In this typical medium speed graph, the blue line corresponds to the E10 fuel, while the red line corresponds to the E0 fuel. Two vertical lines were placed in the graph based on the highest heat release rate point for each fuel. The heat release rate for E0 and E10 was found to be 127.197 and 118.569 KJ/kg deg. at crank angle positions of 14.9° and 14.4°, respectively. It can be seen that at medium speed, the HRR was higher for E0 when compared to E10 by around 6.7%. However, it can also be observed that the maximum heat was released earlier for E10 than E0, which means that the rate of combustion is better for E10 fuel. Table 7 shows a comparison of the combustion pressure and heat release for E0 and E10 at high speed based on the graph shown in Figure 4. In this typical high-speed graph, the blue line corresponds to the E10 fuel, while the red line corresponds to the E0 fuel. Two vertical lines are placed in the graph based on the highest heat release rate point for each individual fuel.



Heat release rate comparison for E0 and E10 at high speeds.

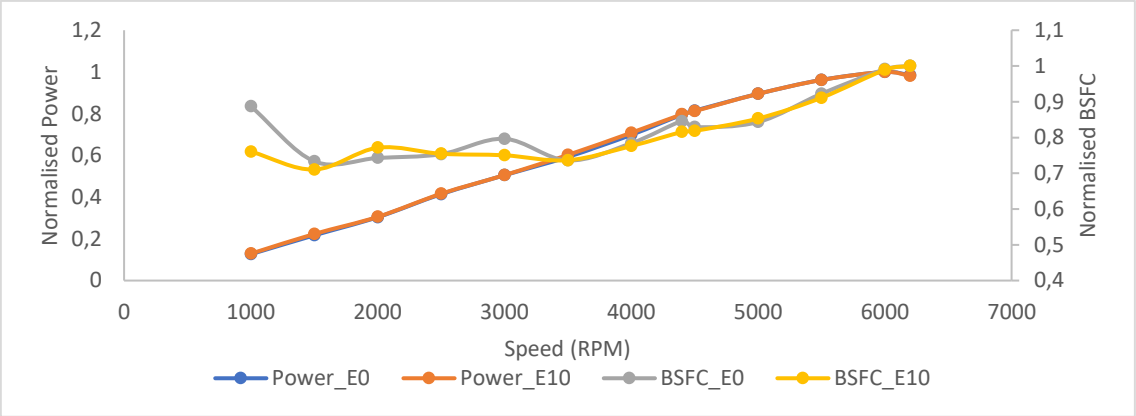
The heat release rate for E0 and E10 was found to be **114.502** and **129.248** KJ/kg deg. at crank angle positions of 14.0° and 12.3°, respectively. In contrast to the low and medium speed results, it can be seen that at high speed the HRR was more for E10 when compared to E0 by around 12.8%. At the same time, it can be observed that the maximum heat was released earlier for E10 than for E0, which means that the rate of combustion is better for the E10 fuel. It can be stated that as the engine speed increases, due to the increase in engine cylinder temperature and homogenous mixture formation, the combustion quality of E10 improves.

Table 7 Combustion pressure and heat release comparison at high speed.

High Speed Category	Crank Angle (Degree)	Pcyl (bar) E0 (Red line)	Pcyl (bar) E10 (Blue line)	dQ_1 (KJ/kg deg) E0 (Red line)	dQ_1 (KJ/kg deg) E10 (Blue line)
Fuel - E0	14.0	**59.323**	**61.803**	**114.502**	**128.843**
Fuel - E10	12.3	**57.111**	**59.477**	**112.455**	**129.248**

Power

The power output given by the engine had a marginal difference among the fuels employed at different rpms for the experimental procedure. The engine performed better with E10 fuel, with an average increase in power of about 1.64% due to better combustion properties. The variation of power is depicted in Figure 5. The rise in power output can be correlated with the HRR graphs for low speed, medium, and high speed in the previous section.



Power and BSFC comparison of averaged-out data for the E0 and E10 fuels.

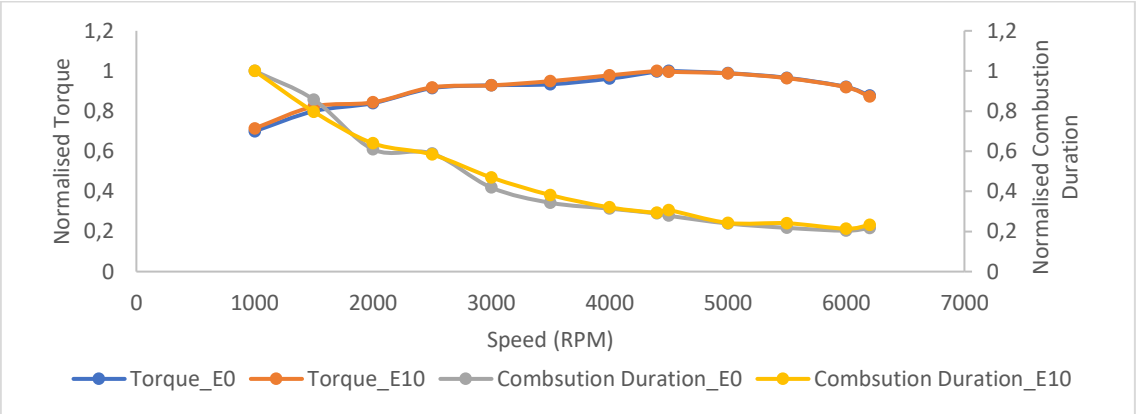
It was observed that HRR was improved for the E10 fuel at higher engine speed and the same resulted in an increase in power with E10 fuel for some of the operating speeds. The main reason behind this could be that the higher-octane number and the inherently present oxygen in the E10 fuel help to improve the combustion quality.

Brake Specific Fuel Consumption

Comparatively, the brake-specific fuel consumption (BSFC) improved again at lower end speeds, pertaining to the better combustion characteristics of E10 at lower end speeds. This trend went up and down due to ECU being calibrated differently for both fuels pertaining to verification for the best performance of the engine. The overall absolute percentage variation in the BSFC was about 3%. This is also depicted in Figure 5. The BSFC is calculated by dividing the fuel flow rate (g/hr) by power (kW) at a particular rpm. As explained above, as the power increased due to E10, accordingly the BSFC values decreased. This means that with the same fuel quantity, more power is produced. The enhancement in combustion quality aids in the fuel consumption improvement with E10 fuel when compared with E0.

Torque

The torque output for E10 increased at all speeds with a 1.74% overall increase in torque output compared to E0 as the heat to work conversion is better, which is also in line with the literature (Eyidogan, *et al.* [29]). This is depicted in Figure 6. The torque is basically a measurement of the engine's ability to perform work. Torque and power are in direct correlation. Since it was observed that the power output was improved at certain operating points for the E10 fuel, accordingly the torque was also improved. Therefore, it can be stated that with E10 fuel, the engine's ability to perform work is improved.



Torque and combustion duration comparison of averaged-out data for the E0 and E10 fuels.

DOI: 10.5614/j.eng.technol.sci.2023.55.3.10

Combustion Duration

Precision in combustion duration measurement is very important in order to understand the dynamics involved in the fuel combustion process. The micro and nano domain seem to intensify and precisely capture the engineering prospects of various processes. Thus, the combustion duration was converted to precise metrics to better differentiate between the fuel combustion levels. The combustion duration was found to be mostly decreasing for E10 compared to E0 by 2.42%, which could be due to the higher laminar flame speed of ethanol [28]. This is also shown in Figure 6. The low combustion duration for E10 means that less time is required for combustion, which in turns suggests that the combustion takes place at a faster rate with the fuel energy released in a shorter span of time. Thus, the combustion quality and in turn the efficiency of the engine is improved with E10 fuel.

TOPSIS methodology

Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is an optimization strategy employed in the design of experiments to come up with a ranking method to transmute the best experimental run based on required output attributes. This process is computationally strong and also emphasizes the domain specific expertise. It caters for the best experimental runs based on Euclidean distance closest to the positive ideal solution and farthest from the negative ideal solution [30]. In our study, we employed this methodology to consider engine performance parameters, which were Power, BSFC, Torque, Mechanical Efficiency. The steps involved in TOPSIS are as follows:

Step 1: The first step involves bifurcating the decision matrix.

Step 2: This step involves normalization of the decision matrix for uniformity of computation. We employed vector normalization, which produces a normalized decision matrix. The vector normalization is given in Eq. (1) as follows:

$$a_{ij}^{*} = \frac{a_{ij}}{\sqrt{\sum_{j=1}^{n} a_{ij}^{2}}} \; for\; j = 1\; to\; n \tag{1}$$

where a_{ij}^{*} represents the vector normalized value for each row (for $1 \le i \le n$) and for each attribute (for $1 \le j \le n$).

Step 3: The normalized decision matrix is then computed by using weighting criteria with the respective weights multiplied by the respective normalized values. This produces the weighted normalized matrix. Simo's weighting criteria were employed and the respective weights for the attributes are given in Table 8. The weighted normalized matrix is given by Eq. (2):

$$wa_{ij}^{*} = a_{ij}^{*} \times w_{j} \tag{2}$$

where, w_j is the weighted value for each attribute, $*$ is the weighted value of the row (for $1 \le i \le n$) and for each attribute (for $1 \le j \le n$).

Table 8 Simo's weighting criterium for performance attributes.

Attributes	Weights
Power	0.3
BSFC	0.3
Torque	0.3
Mechanical Efficiency	0.1

Step 4: The weighted normalized decision matrix is then utilized for computing the Euclidean distance for closeness to the ideal solution. The positive ideal solution (PIS) and the negative ideal solution (NIS) are given by:

$$S_{ij}^{+} = \sqrt{\sum_{i=1}^{n} (wa_{ij}^{*} - p_{j}^{+})^{2}} \;\; \text{for i = 1 to n} \tag{3}$$

$$S_{ij}^{-} = \sqrt{\sum_{i=1}^{n} (wa_{ij}^{*} - n_{j}^{+})^{2}} \;\; \text{for i = 1 to n} \tag{4}$$

where:

S_{ij}^{+} represents the distance closeness to the positive ideal solution

S_{ij}^{-} represents the distance closeness to the negative ideal solution

p_j^{+} represents the best value for each attribute (for $1 \le i \le n$)

n_j^{+} represents the worst value for each attribute (for $1 \le j \le n$)

Step 5: The closeness coefficient is computed and ranked to find out alternatives close to the best solution.

$$\boldsymbol{P_j} = \frac{s_{ij}^{-}}{s_{ij}^{-}+s_{ij}^{+}} \tag{5}$$

where P_j represents the closeness coefficient (TOPSIS Score) and S^{+}, SS^{-} represent the PIS and NIS closeness values.

The workflow of this ranked based methodology involves various steps, i.e., decision matrix formulation, normalized decision matrix, weighted normalized decision matrix, computing the Euclidean distance for PIS and NIS, and closeness coefficient formulation. The TOPSIS procedure was automated in Python according to the aforementioned steps and the ranking hierarchy for the experimental runs for both fuels, E0 and E10, are shown in Tables 9 and 10 respectively.

Table 9 TOPSIS methodology for E0 test data.

Speed Average	W Power	W BSFC	W Torque	W ME %	PIS	NIS	TOPSIS Score	Rank
1000	0.015	0.088	0.064	0.029	0.108	0.011	0.096	13
1500	0.026	0.073	0.073	0.029	0.095	0.03	0.241	12
2000	0.036	0.074	0.076	0.028	0.084	0.035	0.297	11
2500	0.049	0.075	0.083	0.028	0.07	0.046	0.398	10
3000	0.06	0.079	0.085	0.028	0.059	0.053	0.474	9
3500	0.07	0.073	0.085	0.028	0.049	0.065	0.571	8
4000	0.083	0.078	0.088	0.028	0.036	0.075	0.674	7
4400	0.094	0.084	0.091	0.028	0.027	0.085	0.761	6
4500	0.096	0.082	0.091	0.028	0.024	0.088	0.785	4
5000	**0.106**	**0.083**	**0.09**	**0.027**	**0.017**	**0.096**	**0.853**	**1**
5500	0.114	0.091	0.088	0.027	0.02	0.102	0.839	2
6000	0.118	0.098	0.084	0.026	0.027	0.105	0.798	3
6200	0.117	0.099	0.08	0.026	0.029	0.103	0.781	5

Here, the W attribute denotes the weighted normalized decision matrix for each attribute, PIS denotes the positive ideal solution (Euclidean distance closest to the ideal solution) and NIS denotes the negative ideal solution (Euclidean distance farthest from the ideal solution)

Table 10 TOPSIS methodology for E10 test data.

Speed_ Average	W_ Power	W_ BSFC	W Torque	W ME %	PIS	NIS	TOPSIS Score	Rank
1000	0.015	0.077	0.065	0.028	0.106	0.024	0.186	13
1500	0.026	0.072	0.075	0.029	0.093	0.033	0.261	12
2000	0.036	0.078	0.076	0.029	0.084	0.033	0.285	11
2500	0.049	0.076	0.083	0.028	0.07	0.046	0.398	10
3000	0.06	0.076	0.084	0.028	0.059	0.055	0.481	9
3500	0.071	0.074	0.086	0.028	0.047	0.066	0.58	8
4000	0.084	0.078	0.089	0.028	0.035	0.076	0.684	7
4400	0.094	0.082	0.091	0.028	0.026	0.085	0.764	5
4500	0.096	0.083	0.09	0.028	0.025	0.087	0.778	4
5000	**0.106**	**0.086**	**0.09**	**0.027**	**0.019**	**0.095**	**0.833**	**1**
5500	0.114	0.092	0.087	0.027	0.021	0.101	0.828	2
6000	0.118	0.1	0.083	0.026	0.029	0.105	0.782	3
6200	0.116	0.101	0.079	0.026	0.032	0.102	0.763	6

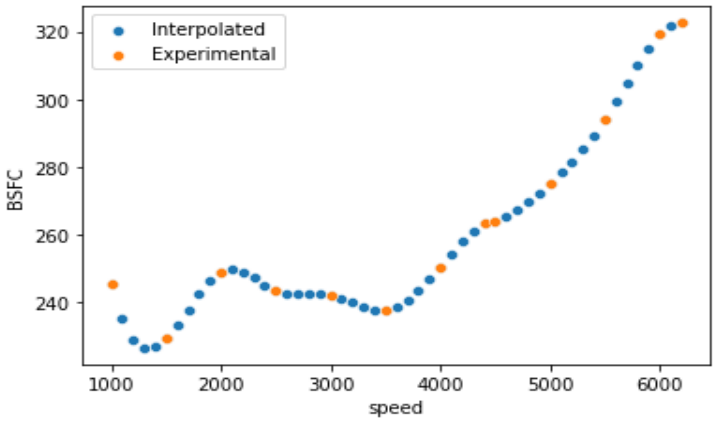
DOI: 10.5614/j.eng.technol.sci.2023.55.3.10

Machine Learning Methodology

Data Preparation

The focus of this study was to investigate the credibility of machine learning methodologies on engine experimental data for different blends of gasoline-ethanol fuels. Our experimental data consisted of only 26 data points, which would not be fruitful for training a machine learning model. The concept of synthetic data generation is utilized in industries to make better models and strive for improvement. This concept has huge potential in the big data era to aid IoT and its superlatives to achieve the concept of digital twins and cyber physical systems on a large scale [23].

On a simpler note, we employed the mathematical methodology of non-linear interpolation to generate supportive synthetic data and converted the 26 experimental runs to 106 experimental data points to make the model training more precise and the predictions more accurate. This non-linear interpolation not only reduced the overall test cost but also the testing time. Figure 7 shows that data were interpolated for both the E0 experimental data and the E10 experimental data. The overall data consisting of 106 experimental runs was used as input for model training using the machine learning methodology.



Interpolated engine performance data to conduct synthetic data generation.

Application of Machine Learning

Ensemble learning methodologies in machine learning have been found to be very effective in making predictions for classification and regression problems. This method includes the aggregation of several decision trees and the corresponding metrics to produce the final decision value. The metrics involved generally include mean squared error, mean absolute error, etc. for regression problems and entropy, Gini index, etc. for classification problems [31]. The workflow for using machine learning in our study considered the speed and fuel properties as input and predicted the engine performance parameters. We used the random forest methodology available in Python from the standard and powerful scikit-learn library, which contain of inbuilt machine learning algorithms. An 80:20 train-test split of data was found to give better results based on trial-and-error observation.

R^2 and mean absolute percentage error (MAPE) were used as metric to check the credibility of the model predictions on the individual output attributes. The R^2 metric corresponds to validating the variance between the actual and the predicted data and MAPE corresponds to the relative error of prediction in actual vs predicted output.

The formulas for MAPE and R^2 are given as below:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|P_i - A_i|}{A_i} \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(P_i - A_i)^2}{\sum_{i=1}^{n}(A_i - \mu)^2} \quad (7)$$

where:

P_i is the predicted value of the i^{th} observation

A_i : Actual value of the i^{th} observation

μ : Mean value of the actual observations

N : number of observations

The result of the prediction vs actual output is tabulated in Table 11.

Table 11 MAPE and R^2 scores of individual performance attributes on experimental data.

Attributes	R^2 score	MAPE (%)
Power	0.997	2.16
BSFC	0.948	1.61
Torque	0.968	1.16
Combustion Duration	0.995	2.62

Based on the result obtained after model training it was found that the R^2 scores for all the attributes were between 0.95 and 0.99; the MAPE (error range) score was found to be between 1% and 3%, which shows the credibility of the ensemble learning algorithm in forecasting of engine performance.

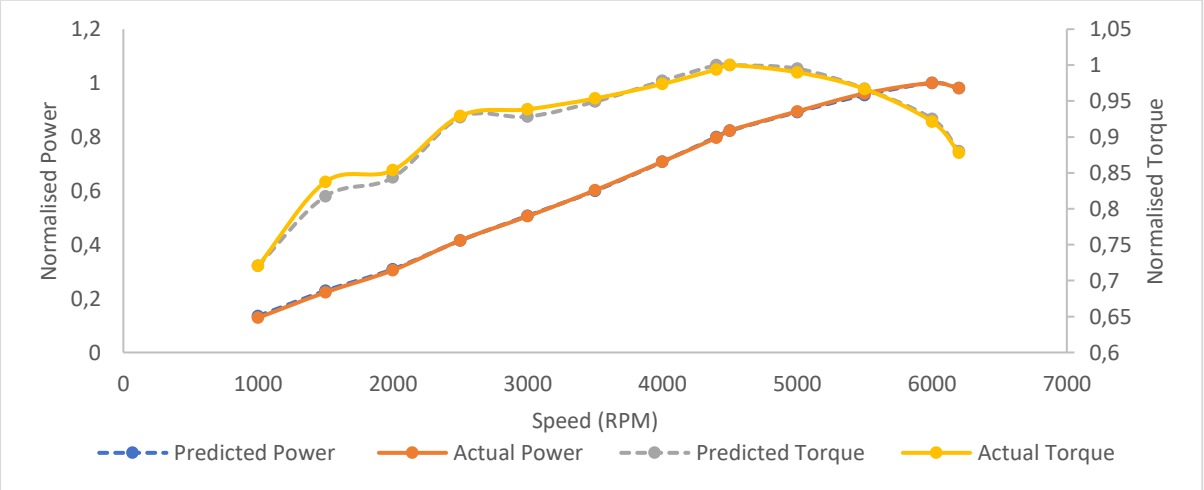
Model Validation

The above validation was done on conventional test data. The model that was trained on the E0 and E10 data was also validated for E15 fuel with respect to the experimental data. The respective performance metrics for the validated data are given in Table 12. The validation for E15 seems to be giving good results, as the model was able to predict all the parameters with good precision. The R^2 score for all the attributes was between 0.95 and 0.99; the MAPE (error range) score was found to be between 1% and 5%.

Table 12 MAPE and R^2 scores of individual performance attributes on E15 test data.

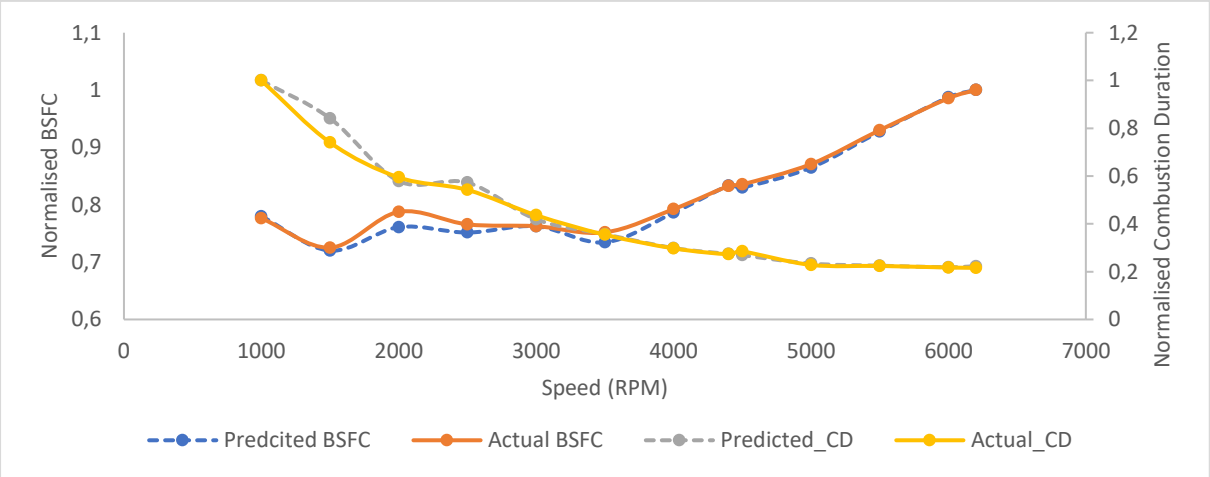
Attributes	R^2 score	MAPE (%)
Power	0.995	3.31
BSFC	0.956	1.76
Torque	0.966	1.38
Combustion Duration	0.980	4.71

Visualization of the prediction vs experimental output is shown for engine power and torque in Figure 8 and for combustion duration (CD) and BSFC in Figure 9.



Predicted vs actual interpretation of engine power and torque for E15 data.

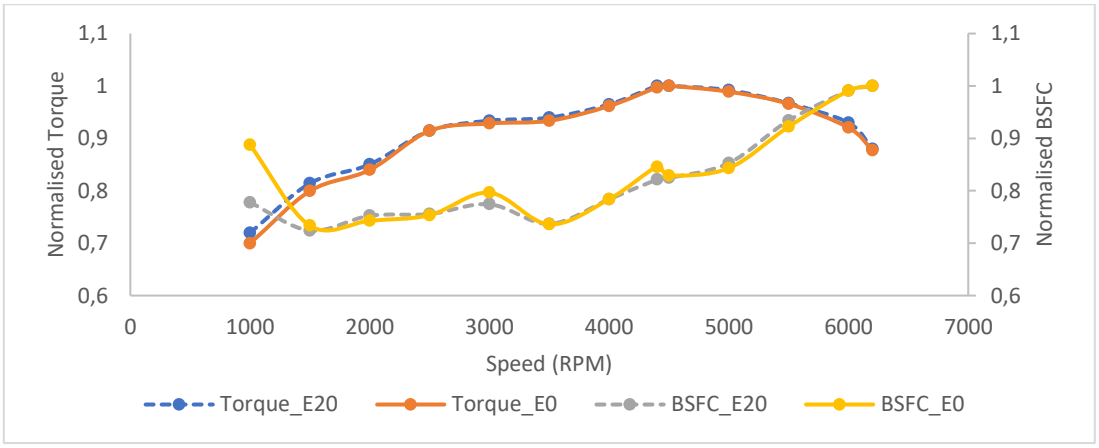
DOI: 10.5614/j.eng.technol.sci.2023.55.3.10



Predicted vs actual interpretation of engine BSFC and combustion duration for E15 data.

Model Prediction

The above model that was trained on the E0 and E10 data was also validated for E15 fuel on the same engine with respect to the experimental data. The predictions were also done on the same engine for E20 fuel and were compared to the E0 actual data. The results are shown in Figure 10.



Predicted E20 vs actual E0 interpretation of engine power and torque.

The torque was on the higher side compared to E0, which could be because oxygenated fuels were used for the lower sets of blends. As the blend percentage increases, the trend will decrease, as the calorific value will substantially overpower the combustion characteristics of oxygenated fuels. BSFC illustrates the mix trend and was found to be higher at higher speeds.

In the same way, the various set of blends can also be predicted and validated based on the study requirements. The importance of TOPSIS along with machine learning could be useful for reducing the time needed for conducting a test whenever a different set of fuels must be tested for performance on the same engine. This can be understood by visualizing the same when predicting the performance of E30 fuel using the same engine and employ TOPSIS to get the ranking based on closeness to the ideal solution. Once the ranking of the experiments is tabulated, decision making is crucial to check for the highest ranked experimental runs or the lowest ranked experimental runs based on the bottlenecks one has to verify. This validation will reduce the overall time and cost involved in any new R&D project in this domain.

Results Comparison with Contemporary Studies

This section presents a brief comparative account of the results obtained by contemporary studies similar to the present work. Sebayang *et al.* [32] applied extreme machine learning models to forecast engine performance of gasoline-bioethanol mix fuels. The authors reported a forecasting error and accuracy (R^2) of less than 3% and above 0.980 respectively. Shin *et al.* [33] applied the Bayesian optimization methodology for predicting diesel engine performance and obtained an accuracy of 0.9675 R^2 with a mean absolute error of 1.6%. Elumalai *et al.* [34] obtained R^2 values of above 0.99 for the estimation of brake thermal efficiency (BTE), brake specific energy consumption (BSEC), and other parameters in a bio-fuelled engine. The authors reported mean absolute percentage errors (MAPE) of 1.12 and 0.84 for BTE and BSEC. The authors concluded that R^2 scores close to unity and errors less than 5% are indicators of good prediction accuracies. As per Yang *et al.* [35], the artificial neural network (ANN) model outperformed the RF model in predicting spark ignition engine combustion. Uslu and Celik [4] reported an (R^2) above 0.964 and a mean relative error (MRE) less than from 0.51% by employing ANN to predict single cylinder diesel engine performance. Considering the above-mentioned recent studies, it is evident that a regression coefficient (R^2) more than 95% and a MAPE less than 5% gives good correlation with the test data and future predictions. In the present study, the values of (R^2) and MAPE were also found to be above 0.95 and less than 5%, respectively, which is in line with the estimation indices followed and prescribed by recent studies in this field.

Conclusions

The literature survey showed the potential of machine learning methods in IC engine application for their ability to map physical processes with intricate detail. Therefore, this study focused on checking for the validation of machine learning models on engine test data. In consideration of flex fuel induction at a large scale around the globe, this process of data driven modelling may help to consider the performance and emissions to identify the bottle necks of vehicles with relative ease. The experimental runs in this study involved test runs on E0 and E10 blends to get suitable performance data for the ML model to be trained upon. After analyzing the details of the comparative data, we found the following:

1. Cylinder pressure was found to be improved for E10 pertaining to its RON number improvement and higher laminar speed compared to E0.
2. Power and torque were found to be increased by 1.64% and 1.74% for E10, which may also correspond to better combustion due to the presence of oxygen in the fuel.
3. BSFC was found to change in an abrupt manner with a variation of about 3%.
4. Combustion duration was found to decrease for E10 by 2.42% due to its higher laminar speeds.

The study also employed the TOPSIS methodology to prepare for the optimized experimental runs based on the engine performance criteria. The 5,000-rpm experiment was found to be the best out of all the experimental data for both fuels, which showed that the engine was running optimally at this speed. The experimental data collected in the present work was synthetically increased from 26 to 106 experimental runs by employing non-linear interpolation. This data was used as input to the ensemble learning method to train the ML model for predicting the performance of the engine considered in the current work. R^2 score and MAPE were used as metrics to validate the ML model, where an R^2 score greater than 0.9 and a MAPE (error range) below 1% to 3% were obtained for all engine performance attributes. This result shows that ML methodologies can be very useful to map the physical processes of engine performance when strongly backed by relevant data. The same ML model was then employed to validate the performance for E15 fuel, which was also compared to actual experimental runs carried out on the same engine. In this case, the R^2 score was again found to be greater than 0.95 and the MAPE (error range) was found to be between 1% and 5% for all the attributes, which indicates that ML modelling can be very useful when several blends of gasoline-ethanol fuels are considered for testing on the same engine. The same ML model was again employed to predict the data for E20 fuel and was compared with the actual E0 data, which also yielded satisfactory results. Hence, ML models can help in making scientifically driven decisions for higher blend engine calibration. The main outcome of this ML based engine performance prediction methodology is the reduction of testing time and cost, while also pre-empting bottle necks for better working efficiency of gasoline-blend fuel engines in the long run.

DOI: 10.5614/j.eng.technol.sci.2023.55.3.10

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

[1] Nguyen, H., Phan, T., Ta, P., Nguyen, C., Pham, N. & Huynh, H., *Gene Family Abundance Visualization based on Feature Selection Combined Deep Learning to Improve Disease Diagnosis*, Journal of Engineering and Technological Sciences, **53** (1), 210109, Jan. 2021. doi: 10.5614/j.eng.technol.sci.2021.53.1.9.

[2] Abdul A.F., Alsaeed M., Sulaiman, S. Mohd A., Mohd K.A. & Al-Hakim, M., *Mixed Reality Improves Education and Training in Assembly Processes,* Journal of Engineering and Technological Sciences, **52** (4), pp. 598, Jul. 2020. doi: 10.5614/j.eng.technol.sci.2020.52.4.10.

[3] Zhou, L., Song, Y., Ji, W. & Wei, H., *Machine Learning for Combustion*, Energy and AI, **7,** pp. 100128, Nov. 2021. doi: 10.1016/j.egyai.2021.100128.

[4] Uslu, S. & Celik, M., *Prediction of Engine Emissions and Performance with Artificial Neural Networks in a Single Cylinder Diesel Engine using Diethyl Ether*, Engineering Science and Technology- An International Journal, **21**, Sep. 2018. doi: 10.1016/j.jestch.2018.08.017.

[5] Sekhar, R., Shah, P., Solke, N. & Singh, T., *Machine Learning-Based Predictive Modeling and Control of Lean Manufacturing in Automotive Parts Manufacturing Industry*, Global Journal of Flexible Systems Management, **23**(1), Oct. 2021. doi:.1007/s40171-021-00291-9.

[6] Jatti, V., Sekhar, R. & Shah, P., *Machine Learning Based Predictive Modeling of Ball Nose End Milling using Exogeneous Autoregressive Moving Average Approach*, IEEE 12th International Conference on Mechanical and Intelligent Manufacturing Technologies, pp. 68-72, May 2021. doi: 1109/ICMIMT52186.2021.9476067.

[7] Afzal, A., Aabid, A., Khan, A., Khan, S., Rajak, U., Verma, T. & Kumar, R., *Response Surface Analysis, Clustering, and Random Forest Regression of Pressure in suddenly Expanded High-Speed Aerodynamic Flows*, Aerospace Science and Technology, **107**, Oct. 2020. doi: 10.1016/j.ast.2020.106318.

[8] Mokashi, I., Afzal, A., Khan, S., Abdullah, N., Azami, M., Jilte, R. & Samuel, O., *Nusselt Number Analysis from a Battery Pack Cooled by Different Fluids and Multiple Back-propagation Modelling using Feed-Forward Networks*. International Journal of Thermal Sciences, **161**(23), Nov. 2020. doi: 10.1016/j.ijthermalsci.2020.106738.

[9] Kodancha, P., Pai, A., Kini, C. & Bayar, R., *Performance Evaluation of Homogeneous Charge Compression Ignition Combustion Engine – A Review*, Journal of Engineering and Technological Sciences, **52**(3), pp. 289-309, May 2020. doi: 10.5614/j.eng.technol.sci.2020.52.3.1.

[10] Singh, G., Dogra, D., Ramana, R., Chawla, J., Sutar, P.S., Sagare, V.S., Sonawane, S.B., Kavathekar, K., Rairikar, S. & Thipse, S.S., *Development of Dual Fuel (Diesel + CNG) Engine for Off-Road Application*, SAE Technical Paper 2021-26-0119, Sep. 2021. doi: 10.4271/2021-26-0119.

[11] Bawase, M.A. & Thipse, D.S.S., *Impact of 20% Ethanol-blended Gasoline (E20) on Metals and Non-metals used in Fuel-system Components of Vehicles*, ARAI Journal of Mobility Technology, **1**(1), pp 1-8, Dec. 2021. doi: 10.37285/ajmt.1.0.1.

[12] Kavathekar, K., Thipse, S., Rairikar, S., Sonawane, S., Sutar, P. & Bandyopadhyay, D., *Study of Effect on Engine Performance Using 15% HCNG Blend Versus CNG Using a Simulation Approach*, Advances in Mechanical Engineering, eds. Kalamkar, V. & Monkova, K. (Springer), Jan. 2021. doi: 1007/978-981-15-3639-7.

[13] Bandyopadhyay, D., Sutar, P., Sonawane, S. & Rairikar, S., *Experimental Analysis of Heavy Duty CNG Engine Based on Its Aspiration and Fuel System*, SAE Technical Paper, pp. 2021–2047, Sep. 2021. doi: 10.4271/2021-26-0117.

[14] Yesilyurt, M., Uslu, S., & Yaman, H., *Modeling of a Port Fuel Injection Spark-Ignition Engine with Different Compression Ratios Using Methanol Blends with the Response Surface Methodology,* Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering, Jul. 2022. doi: 10.1177/09544089221112373.

[15] Badra, J., Khaled, F., Sim, J. & Pei, Y., *Combustion System Optimization of a Light-Duty GCI Engine Using CFD and Machine Learning*, SAE Technical Paper 2020-01-1313, Apr. 2020. doi: 10. 4271/2020-01-1313.

[16] González, S., Kroyan, Y., Sarjovaara, T., Kiiski, U., Karvo, A., Toldy, A., Larmi, M., & Santasalo-Aarnio, A., *Prediction of Gasoline Blend Ignition Characteristics Using Machine Learning Models,* Energy & Fuels, **35** (11), pp. 9332–9340, May 2021. doi: 10.1021/acs.energyfuels.1c00749.

[17] Airamadan, A., Ibrahim, Z., Mohan, B., & Badra, J., *Machine Learning Model for Spark-Assisted Gasoline Compression Ignition Engine,* SAE International Journal of Advances and Current Practices in Mobility, **5** (2), Mar. 2022. doi: 10.4271/2022-01-0459.

[18] Shah, N., Zhao, P., Delvescovo, D. & Ge, H., *Prediction of Autoignition and Flame Properties for Multicomponent Fuels Using Machine Learning Techniques,* SAE Technical Paper, pp. 2019–2020, Mar. 2019. doi: 10.4271/2019-01-1049.

[19] Petrucci, L., Ricci, F., Mariani, F. & Cruccolini, V., *Engine Knock Evaluation Using a Machine Learning Approach*, SAE *Conference on Sustainable Mobility,* SAE Technical Paper 2020-24-0005, Sep. 2020. doi: 10.4271/2020-24-0005

[20] Chandrasekar, K., Sudhakar, S., Rajappan, R., Senthil, S. & Balu, P., *Present Developments and the Reach of Alternative Fuel: A review*, Materials Today Proceedings, **51**(3), May 2021. doi: 10.1016/j.matpr.2021.04.505.

[21] Bielaczyc, P., Woodburn, J., Klimkiewicz, D., Pajdowski, P., & Szczotka, A., *An examination of the effect of ethanol–gasoline blends' physicochemical properties on emissions from a light-duty spark ignition engine*, Fuel Processing Technology, **107**(5813), pp. 50–63, 2013. doi: 10.1016/j.fuproc.2012.07.030.

[22] Mohamad, B., Szepesi, G.L., & Bollo, B., *Review Article: Effect of Ethanol-Gasoline Fuel Blends on the Exhaust Emissions and Characteristics of SI Engines*, Vehicle and Automotive Engineering, **2**, VAE 2018, May 2018, Lecture Notes in Mechanical Engineering. Springer, Cham. doi: 10.1007/978-3-319-75677-6_3.

[23] He, Z. & Zhou, W., *Generation of synthetic full-scale burst test data for corroded pipelines using the tabular generative adversarial network*, Engineering Applications of Artificial Intelligence, **115**(7), Oct. 2022. doi: 105308. 10.1016/j.engappai.2022.105308.

[24] Badra J., Owoyele O., Pal P., & Sibendu S., *A machine learning-genetic algorithm approach for rapid optimization of internal combustion engines*, Artificial Intelligence and Data Driven Optimization of Internal Combustion Engines, ed.1, Elsevier, pp. 125-158, Jan. 2022. doi: 10.1016/B978-0-323-88457-0.00003-5.

[25] Mariani, V., & Och, S., & Coelho, L., & Domingues, E., *Pressure prediction of a spark ignition single cylinder engine using optimized extreme learning machine models*, Applied Energy, pp. 249:204-221, Sep. 2019. doi: 10.1016/j.apenergy.2019.04.126.

[26] Togun, N., & Baysec, S., *Prediction of torque and specific fuel consumption of a gasoline engine by using artificial neural networks*. Applied Energy, **87**(1), pp. 349-355, Jan. 2020. doi: 10.1016/j.apenergy.2009.08.016.

[27] Najafi, G., Ghobadian, B., Tavakoli, T., Buttsworth, D., Yusaf, T.F., & Faizollahnejad, M., *Performance and exhaust emissions of a gasoline engine with ethanol blended gasoline fuels using artificial neural network*, Applied Energy, **86**(5), pp. 630-639, May 2009. doi: 10.1016/j.apenergy.2008.09.017.

[28] Baddu, N., Khalid, A., Samsudin, D., Zaman, I., & Manshoor, B., *Investigation of Flame Characteristics of Ethanol-Gasoline Blends Combustion Using Constant Volume Chamber*, MATEC Web of Conferences, **78,** 01030, Jan. 2016. doi: 10.1051/matecconf/20167801030.

[29] Eyidogan, M., Ozsezen, A. N., Canakci, M. & Turkcan, A*., Impact of alcohol-gasoline fuel blends on the performance and combustion characteristics of an SI engine*, Fuel*,* **89** (10), pp. 2713–2720, Oct. 2010. doi: 10.1016/j.fuel.2010.01.032.

[30] Sakthivel, G., Ilangkumaran, M., & Gaikwad, A., *A Hybrid Multi-Criteria Decision Modelling Approach for The Best Biodiesel Blend Selection Based On ANP-TOPSIS Analysis*, Ain Shams Engineering Journal, **6**(1), pp. 239–256, Sep. 2014. doi: 10.1016/j.asej.2014.08.003.

[31] Khoshkangini, R., Mashhadi, P., Tegnered, D., Lundstrom, J., & Rognvaldsson, T., *Predicting Vehicle Behavior Using Multi-Task Ensemble Learning*, Expert Systems with Applications*,* **212**(6), Sep. 2022. doi: 10.1016/j.eswa.2022.118716.

[32] Sebayang, A., Masjuki, H., Ong, H. C., Dharma, S., Silitonga, A., & Kusumo, F., *Prediction of Engine Performance and Emissions with Manihot Glaziovii Bioethanol- Gasoline Blended Using Extreme Learning Machine*, Fuel, **210** (7), pp. 914–921, Sep. 2017. doi: 10.1016/j.fuel.2017.08.102.

[33] Shin, S., Lee, Y., Kim, M., Park, J., Lee, S., & Min, K., *Deep neural network model with Bayesian hyperparameter optimization for prediction of NOx at transient conditions in a diesel engine*. Engineering Applications of Artificial Intelligence, **94** (Feb), Sep. 2020. doi: 103761. 10.1016/j.engappai.2020.103761.

DOI: 10.5614/j.eng.technol.sci.2023.55.3.10

[34] Elumalai, P. V., Krishna M., R., Parthasarathy, M., Samuel, O. D., Owamah, H. I. & Saleel, C. A., *Artificial Neural Networks Model for Predicting the Behavior of Different Injection Pressure Characteristics Powered by Blend of Biofuel-Nano Emulsion*, Energy Science & Engineering, **10** (7), pp. 2367–2396, Apr. 2022. doi: 10.1002/ese3.1144.

[35] Yang, R., Yan, Y., Sijia, R. & Liu, Z., *Modeling Performance and Emissions of a Spark Ignition Engine with Machine Learning Approaches*, SAE International Journal of Advances and Current Practices in Mobility, SAE Technical Paper 2022-01-0380, **5**(2), Mar. 2022. doi: 10.4271/ 2022-01-0380.